

RECEIVED
STIC
BIOTECH/ CHEMICALU.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE

ONLINE SEARCH REQUEST FORM

USER

Spector

U.S. SERIAL NUMBER

1993 SEP 27 AM 10:54 252491

ART UNIT

1812

PHONE

308-1793

DATE

9/27/95

Please give a detailed statement of requirements. Describe as specifically as possible the subject matter to be searched. Define any terms that may have special meaning. Give examples or relevant citations, authors, or keywords, if known.

You may include a copy of the broadest and/or relevant claim(s).

*Please search Seq ID No: 1, 2,
4, 18*

FILE

FILE

FILE

FILE

FILE

STAFF USE ONLY

COMPLETED

9-28-95

SEARCHER

SILIP X308 4268

ONLINE TIME

20

TOTAL TIME

25

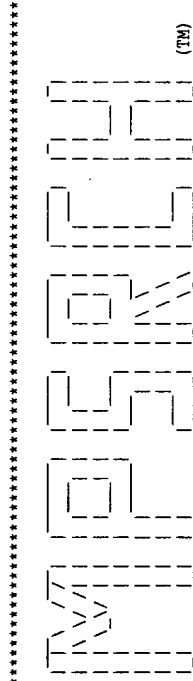
(in minutes)

NO. OF DATABASES

11

SYSTEMS

☐ CAS ONLINE☐ DARC/QUESTEL☐ DIALOG☐ SDC☒ OTHER *MPSRCU*



Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPearch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Sep 28 00:03:50 1995; MacPar time 84.88 Seconds
Tabular output not generated. 893.421 Million cell updates/sec

Title: >US-08-252-491-1
Description: (1:1486) from US08252491.seq
Perfect Score: 1486
N.A. Sequence: 1 CTCGTCGGCGTCTCAGGC.....TATTAATTGAAATCACTA 1486
Comp: GGACGACGGCCAGCACTCG.....ATATTAACTTTTAGTGAT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 61539 seqs, 25515148 bases x 2

Database: n-geneseq
1 n-gen1
2 n-gen2
3 n-gen3
4 n-gen4
5 n-gen5
6 n-gen6
7 n-gen7
8 n-gen8
9 n-gen9
10 n-gen10
11 n-gen11

Statistics: Mean 9.598; Variance 5.610; scale 1.711

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	95	6.4	1047	2	Q10572 Human Natriuretic Pep 1.06e-44	
2	71	4.8	1047	2	Q10572 Human Natriuretic Pep 8.33e-29	
3	43	2.9	91	9	Q51746 Oligonucleotide probe 2.56e-11	
4	43	2.9	204	1	N81164 Base substituted E.co 2.56e-11	

c	5	42	2.8	91	9	Q51746 Oligonucleotide probe	9.90e-11
c	6	39	2.6	204	1	N81164 Base substituted E.co	5.46e-09
c	7	32	2.2	565	6	Q35072 HCV envelope region n	4.37e-05
c	8	28	1.9	565	6	Q35072 HCV envelope region n	5.52e-03
c	9	25	1.7	3871	2	N71302 HSV-1 gB and surround	1.72e-01
c	10	24	1.6	1542	3	Q20695 Mouse OSF.1.	5.18e-01
c	11	23	1.5	36	2	Q11195 Ballast Constituent c	1.52e+00
c	12	23	1.5	36	2	Q11195 Ballast Constituent c	1.52e+00
c	13	22	1.5	33	2	Q11197 Ballast Constituent c	4.35e+00
c	14	22	1.5	33	7	Q46603 Mixed oligonucleotide	4.35e+00
c	15	22	1.5	39	7	Q51787 Mixed oligonucleotide	4.35e+00
c	16	22	1.5	39	7	Q51787 Mixed oligonucleotide	4.35e+00
c	17	21	1.4	33	2	Q11197 Ballast Constituent c	1.21e+01
c	18	21	1.4	33	7	Q46603 Mixed oligonucleotide	1.21e+01
c	19	21	1.4	1830	2	N60106 Generic sequence enco	1.21e+01
c	20	21	1.4	2549	1	N91467 Sequence of human BP5	1.21e+01
c	21	21	1.4	12492	1	N70543 Complete sequence of	1.21e+01
c	22	21	1.4	498	3	N50034 Sequence encoding new	1.21e+01
c	23	21	1.4	3871	2	N71302 HSV-1 gB and surround	1.21e+01
c	24	20	1.3	320	8	Q60277 Human brain Expressed	3.26e+01
c	25	20	1.3	1908	2	N71064 Gene encoding Plasmid	3.26e+01
c	26	20	1.3	2086	3	Q23455 Human lactoferrin pro	3.26e+01
c	27	20	1.3	501	3	N50025 Sequence encoding new	3.26e+01
c	28	20	1.3	2159	10	Q67223 Mouse p55Nuc.	3.26e+01
c	29	20	1.3	1494	1	N90344 Plasmid pfc-epsilon-R	3.26e+01
c	30	20	1.3	1305	1	N81652 Alpha-amino-epsilon c	3.26e+01
c	31	20	1.3	1515	1	N90134 Fc epsilon receptor g	3.26e+01
c	32	20	1.3	48	7	Q46605 Mixed oligonucleotide	3.26e+01
c	33	20	1.3	416	4	N10053 Generic coding sequen	3.26e+01
c	34	20	1.3	1504	1	N81367 EcoRI insert from pfc	3.26e+01
c	35	20	1.3	9643	1	N80859 Sequence of entire HI	3.26e+01
c	36	20	1.3	501	8	Q49205 Field hamster Aphrodi	3.26e+01
c	37	20	1.3	4854	5	Q28399 Human nestin gene.	3.26e+01
c	38	20	1.3	1566	1	N82198 GAGRODN sequence from	3.26e+01
c	39	20	1.3	1079	4	Q27199 Human low affinity Fc	3.26e+01
c	40	20	1.3	501	3	N50033 ACC deaminase.	3.26e+01
c	41	20	1.3	501	3	N50033 Sequence encoding new	3.26e+01
c	42	20	1.3	30	7	Q46602 Mixed oligonucleotide	3.26e+01
c	43	20	1.3	747	8	Q49207 Field hamster Aphrodi	3.26e+01
c	44	20	1.3	1504	3	N81512 Sequence of a gene fo	3.26e+01
c	45	20	1.3	9672	3	Q20616 ROD HIV-2 isolate com	3.26e+01

ALIGNMENTS

RESULT	1
ID	Q10572 standard; DNA; 1047 BP.
AC	Q10572;
DT	09-APR-1991 (first entry)
DE	Human Natriuretic Peptide Receptor B.
RW	NPBR; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW	hyperaldosteronism; glaucoma; guanyl cyclase.
OS	Homo sapiens.
FH	Key Location/Qualifiers
FT	Peptide 1..22
FT	/label= signal sequence
FT	Protein 12
FT	/label= mature NPBR
FT	Domain 23..455
FT	/label= extracellular domain
FT	/note= *binds natriuretic peptides A, B and C "
FT	Domain 456..456
FT	/label= transmembrane domain
FT	Domain 479..1047

FT /label= cytoplasmic domain
 - FT /note= "GC and protien kinase activity"
 FT Modified -site 24..26
 FT /label= N-glycos_site
 FT Modified -site 35..37
 FT /label= N-glycos_site
 FT Modified -site 161..163
 FT /label= N-glycos_site
 FT Modified -site 195..197
 FT /label= N-glycos_site
 FT Modified -site 244..246
 FT /label= N-glycos_site
 FT Modified -site 277..279
 FT /label= N-glycos_site
 FT Modified -site 349..351
 FT /label= N-glycos_site
 FT Modified -site 600..602
 FT /label= N-glycos_site
 PN W09100292-A.
 PD 10-JAN-1991.
 PF 22-JUN-1990; U03586.
 PR 23-JUN-1989; US-370673.
 PA (GETH) GENENTECH INC.
 PI Chang M, Goeddel D, Lowe D;
 DR WPI; 91-036711/05.
 DR N-PSDB; Q10324.
 PT Natriuretic protein receptor B - for diagnosis and treatment of
 PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
 PS Claim 3; Fig 1; 49pp; English.
 CC The sequence was derived from the DNA encoding natriuretic peptide
 CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein
 CC kinase activity. The DNA can be inserted into expression vectors
 CC for the prodn. of the protein, opt. after being mutated to produce
 CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
 CC 114,952). The protein (or variants) can be used in treatment of
 CC natriuretic peptide disorders, and also to isolate peptides using
 CC affinity chromatography. Antibodies with affinity for NPRB can
 CC also be prepd.
 SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;
 DB 2; Score 95; Match 8.9%; QryMatch 6.4%; Pred. No. 1.06e-44;
 Matches 88; Conservative 278; Mismatches 615; Indels 9; Gaps 9;
 Db 17 vrnngarntnnavnnhnsyawaarvgnvnavnavnangrannvdrhnsennngacs 76
 Cp 1004 GGGCGAGATGCTGAGGGTTCCAGGGCTCGAGTGGTTCACAGAGATTCGGGACCTGGA 945
 Db 77 nynnnaevdnknyhdnndnngvcvynnaasvarn-ashwrrnnnttagavasgsakndh 135
 Cp 944 TCACATTCACAGGTCGCTGTCTGTCTGATCCAGATTCACAGATTCGGGACCTGGA 885
 Db 136 yrtvrtvqneankngnvvtnhghnwtaraannynndardtdrnhvntnngvnnanngsn 195
 Cp 884 GGTTCATTAGTTCACAGGAGTAATCTTGACTCTGAATCCCTCAAGCCTGCTCAGAAG 825
 Db 196 nsvhnvyarnggnnathnraigrvncngnmmhnnnnnnnnnnnnnnnnnnnnnnnnnn 255
 Cp 824 TCCAGGGCCAGGAGTTCTGGCTGTGACACTGAGTTCGTCTCCCAACATCCAGAGTCT 765
 Db 256 dvngnsnragtnratgrndhrttrnnanranrnanntvntvrrnnnnnnnnnnnnnnnn 315
 Cp 764 GTTGGGACCTGTTAGTGTGAGGAGTTGAGAGTACTGCTTGGGACAGCTGTGGTGG 705
 Db 316 rarndngvngnsnnnnn-agcnydgnnnyanvnnntnnnggrtndgnrnnvnmkgrryh 374

Cp 704 CAGGGTCCCTCTGCACACAGAGGTGGGACCTTCTACCAAGAGCAGGAAGCGCACCTTC 645
 Db 375 gvtgnvmdknnndrntdrvnwmgdndsgdnnaahysganknmwftgrnnnwvkgans 434
 Cp 644 CGAAGCAGTGTGTTGCAAGCTCAAGAAGAGGCA-TTGGGGTCTTGTGAGCTGTGGTCC 586
 Db 435 dnmncandndsdck-tnnstnanvngtgnthnmngvssnnnnrknmmnknmasnwr 493
 Cp 585 TGCCTCTAGAGGAAGTGGTTCCTAGAGGGCCTGCAGGGGCCCAAGAGGAGCGGA 526
 Db 494 nrwnnnngsnryhkgagrntnsnrgssysmntahgkykynnmantghnkvvvanhkv 553
 Cp 525 CCTGCCACGAAAGTGTCCAGAGGGATGAGAGGAGAGGGTTCCAACTGTCTCTGTG 466
 Db 554 nkknrntnrvnnnnkhmrdivnnhnrnngacndnnnncvntvncrigenndnnnden 613
 Cp 465 CTGCCATCATCTCCCTCCAGTAGAGGACACTGCCCTAGAAATGTCTGTGCTTGTCT 406
 Db 614 nndwmrnyssnnndvkgmannhnsnshgsknsncvdsrvnknktdygnasnrstan 673
 Cp 405 GTTCGGTCTGGGTTTTCATTTCCATTTCCAGGCTAAAGTCCACAGAGGAGGAGGGA 346
 Db 674 nddhnanyakkntannnsgnnnttgmaadvysngnnnnnnnnnnnnnnnnnnnnnn 733
 Cp 345 TAGACAAAGGGTCGAGCTGGGACACTGACTCAGTGGCTGTGAAGGAGG-TGGGAGTCA 287
 Db 734 knnvkrngnrynrnsndrtnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnkng 793
 Cp 286 CCGACGAGTTTATTAGAGACTCTGGGCTACAGGAGGAGCTACGGGGCTGCACAGATT 227
 Db 794 tgnndnnnnrnmnyannnkvnrrtnaynnnrkanannnnnnnnhsvaannkrgntvn 853
 Cp 226 A-GCTTGGCCATCGCAAGAGCATGGCCGCGAGGAGCAATCAGTCAGCTCCATT-CTGG 169
 Db 854 anadsvtnynsdvngvntaneasntnmvntnnndnytcndamndndvkvntngdaym 913
 Cp 168 CCAGGTCTTAAGTCAAGTGGCTCCCTGTATGGGGCTCTCCCTCGAATCT-TTCCTGGG 110
 Db 914 vvsngnrgnrgnannanmanndavsnrnrhndmnrnngvhtgnvcagvgnk 973
 Cp 109 GCATGGAGGGGGCTTCAGTCTGTCTTCTTTCGGCAGAGTTCGGGTGGGAGAGGTG 50
 Db 974 mnrvcmgdvtvntasrmnngnnknhsvs 1003
 Cp 49 -GGCCAAGAGCTCTGTCCGGGTGGAGAGG 21

RESULT 2
 ID Q10572 standard; DNA; 1047 BP.
 AC Q10572;
 DT 09-APR-1991 (first entry)
 DE Human Natriuretic Peptide Receptor B.
 KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
 KW hyperaldosteronism; glaucoma; guanylyl cyclase.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= signal sequence
 FT Protein 12
 FT /label= mature NPRB
 FT Domain 23..455
 FT /label= extracellular domain
 FT /note= "binds natriuretic peptides A,B and C]"

[illegible][illegible]

DR	WPI; 88-279927/40.
-	PT Introducing random point mutations into nucleic acids -
PT	by prepn of single stranded template, annealing a primer, elongation,
PT	misincorporation, completion of molecules and screening.
PS	Disclosure; p; English.
-	CC Random point mutations were introduced into the alpha fragment of
CC	E.coli beta-galactosidase. The wild type sequence was obtained as a
CC	single stranded template and an oligonucleotide was hybridised to
CC	it to generate a popn of DNA molecules which terminate at all
CC	possible nucleotide positions within a specified region. The
CC	variable 3' ends generated in this way are used as primers for
CC	reverse transcriptase. Nucleotides are misincorporated by the
CC	transcriptase and the molecules are completed to forms that can be
CC	amplified and then expressed in a suitable host-vector system.
CC	The sequence covers all 176 diffnt base substitutions, most of which
CC	occurred singularly in any given mutant.
CC	See also P8057S.
SQ	Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

DB 1; Score 39; Match 11.2%; QryMatch 2.6%; Pred. No. 5.46e-09; Matches 13; Conservative 60; Mismatches 42; Indels 1; Gaps 1;

Db 76 ycdhvgccgymrttthhyrimrbvnydrnsdaawccyrrsvkydcccynachddh 135
:::|::| | : : : :::: :: ||: : : ::||: | ::::
Cp 1448 TCCTGACTGCTCTGTATGCATATTGCTGCTAGGTAAAAAATAATAGTCCFAAAATTTT 1389

Db 136 -yyvbbvynhnhnncnccbnhvhcnhbnhrnwayvrhdarddvheevch 190
:: :::: :: : : |:: : : :: : :::: :::

Cp 1388 ACAATCTCCAGTGTGTGATCCCTTCCCAGGGCCTTTAGGTGAAAGCAGAACA 1333

RESULT

ID Q35072 standard; DNA; 565 BP.
AC Q35072;
DT 20-MAY-1993 (first entry)
DE HCV envelope region nucleic acid.
KW Envelope; region; type C; hepatitis; virus; HCV; vaccine; serum;
KW non-A, non-B; amplify; ss.
OS Hepatitis C virus.
PN J04349885-A.
PD 04-DEC-1992.
PF 29-MAY-1991; 152169.
PR 29-MAY-1991; JP-152169.
PA (TEIJ) TEIJUN LTD.
PT WP1; 93-022708/03.
PT Envelope region nucleic acid fragment - for type C hepatitis
PT virus (I), for producing vaccine
PS Claim 1; Page 2; 13pp; Japanese.
CC This sequence encodes a novel envelope region of type C hepatitis
CC virus (HCV). This fragment can be used for the preparation of a
CC vaccine for hepatitis C. This fragment was prepared from the serum
CC of non-A, non-B hepatitis patients and the envelope region DNA was
CC amplified by PCR using the primer sequences given in Q35073-76.
SQ Sequence 565 BP; 61 A; 92 C; 106 G; 85 T;

DB 6; Score 32; Match 34.6%; QryMatch 2.2%; Pred. No. 4.37e-05;
Matches 18; Conservative 24; Mismatches 10; Indels 0; Gaps 0;

Db 435 ctmbcmymgmdgyavgdgarrdyrcargrytgaaaytgycdmtayscy 486
|| ::||::: ||::||:: : ||::| ::||: ||: : :::

Cp 512 CTGTCCCAGGAGGATGAGAGGCAGGAGGTTCCAACTGTCTCGTGTGCC 461

RESULT

[illegible]

RESULTS

ID	N71302	standard; DNA; 3871 BP.
AC	N71302;	
DT	30-APR-1991	(first entry)
DE	HSV-1 gB and surrounding regions.	
KW	Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;	
KW	glycoprotein; gB; ss.	
OS	Herpes simplex virus type 1 (KOS).	
FH	Key	Location/Qualifiers
FT	misc_RNA	1..375
FT	/*tag= a	
FT	/note= '5'	extra sequences beginning with the XhoI
FT	site*	
FT	CRAT_signal	406..410
FT	/*tag= b	
FT	/number= 1	
FT	CRAT_signal	443..448
FT	/*tag= c	
FT	/number= 2	
FT	TATA_signal	476..479
FT	/*tag= d	
FT	misc_RNA	501..789
FT	/*tag= e	
FT	/label= mRNA	start sequence
FT	/note= '501 is possible	start site*'

FT misc_RNA 504...789
 FT /*tag= f
 FT /label= mRNA start sequence
 FT /note= "504 is possible start site"
 FT misc_RNA 506..789
 FT /*tag= g
 FT /label= mRNA start sequence
 FT /note= "506 is possible start site"
 CDS 790..3498
 FT /*tag= h
 FT /label= HSV-1 gB
 FT /note= "includes N-terminal hydrophobic leader and
 a membrane-spanning sequence, a C-terminal
 ionic sequence, and 9 N-linked
 FT saccharide-addition sites"
 FT 3'UTR 3499..3549
 FT /*tag= i
 FT polyA_signal 3518..3525
 FT /*tag= j
 FT polyA_site 3549..3549
 FT /*tag= k
 FT misc_RNA 3549..3997
 FT /*tag= l
 FT /note= "3' nonessential sequences to the BamHI
 FT site"
 FT US4642333-A.
 PN 10-FEB-1987.
 PD 20-JUN-1984; 622496.
 PF 16-SEP-1983; US-532996.
 PR 20-JUN-1985; US-622496.
 PA (PERS/) PERSON S.
 PI Person S;
 DR WPI; 87-056354/08.
 PT Amino acid chain of glyco:protein B of HSV-1 and 2 - prepd. as
 recombinant and used for vaccines for herpes simplex virus types 1
 PT and 2.
 PS Example; Table 1; 16pp; English.
 CC !NOTE! This sequence has been indexed as represented in the
 CC specification, except that bases 'E' have been replaced by 'N'.
 CC The features have been indexed according to the legend of table 1 on
 CC column 19/20 and the Sequence Summary of column 5 (sic). Note that
 CC the base numbering of the features does not correspond to the
 CC the sequence numbering below.
 CC For another DNA sequence of HSB-1 gB see N71303 (P71135),
 CC and for HSV-2 gB see N71399 (P71136).
 CC A pure non-glycosylated amino acid (AA) chain comprising a sequence
 CC corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2
 CC virus which is antigenic to HSV-1 of HSV-2, which contains no more
 CC than 750 AA residues, and which includes AA residues 135-649
 CC inclusive is claimed. It can be used to produce vaccines for
 CC prophylaxis and treatment of HSV-1 and HSV-2.
 SQ Sequence 3871 BP; 743 A; 1402 C; 544 G; 754 T;

DB 2; Score 25; Match 39.4%; QryMatch 1.7%; Pred.No. 1.72e-01;
Matches 54; Conservative 3%; Mismatches 49; Indels 2; Gaps 2;

Dn 2936 cbtbbcgctatcbaebtcggbttbtcbctctcatbtcgaacccctttctttbbb 2995
| : :: | | | : | | | : : | : | | | | | | | | | | | | : : | :

Qy 1022 CTCCTGGCATTCACCTCCAGGPGACTTCTCCTTCTCAAGGCCTGTCTCATGG 1081

Dn 2996 bqbcbtgbcbbtbbctbttbbcttbbcgcgtggcbbectttcbbctttcbtta 3055
: : ||||::||::||::||::||::||::||::||::||::||::||::||::||:

Qy 1082 ACACAGCCCTTCCTCGT-CATCA-CCTGCCCTTGCCACCACCATCGATCTCACGCCGAG 1139

```

Db 3056 cbtcatcbcbtcbtcb 3072
    I: ||::||::|:|
Qy 1140 CTCACCCCGCTGTTCC 1156

RESULT 10
ID Q20695 standard; cDNA; 1542 BP.
AC Q20695;
DT DT 24-APR-1992 (first entry)
DE Mouse OSF.1.
KW MK; calvarial cell; probe; diagnosis; osteoblast;
KW cranial nerve cell; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 229..735
FT /*tag= a
FT /label= I-I
PN W09200324-A.
PD PD 09-JAN-1992.
PF 27-JUN-1991; J00871.
PR 29-JUN-1990; JP-169824.
PR 28-SEP-1990; JP-256810.
PA (FARH ) HOECHST JAPAN LTD.
PI Hashimoto T, Tezuka K, Kumegawa S, Takagi C;
DR WPI; 92-041516/05.
DR P-PSDB; R20238.
PT Protein from mouse calvarial cells - differentiates growth of
PT osteoblast(s) and cranial nerve cells, for treating and
PT diagnosing osteoporosis and dementia
PS Disclosure; Fig 3(1-5); 43pp; Japanese.
CC A cDNA library was constructed from mouse calvarial cell line
CC MC3T3E1. Differential screening in NIH3T3 cells (ATCC CRL-1658) and
CC cloning, then insertion into vector pUC118 gave plasmid pMC031 which
CC contains the DNA coding for mouse OSF-1 (Q20695). This plasmid was
CC used as a probe to clone by plaque hybridisation from a human cDNA
CC library to give a phage clone, HBRI. Insertion into vector pUC118
CC gave pHRI1 which may be used diagnostically or the OSF-1 cDNA
CC (Q20696) incorporated (with poly-A signal sequence and SV40 T-cell
CC antigen promoter) into a vector (such as pHSG-757) for expression
CC (e.g. in CHO cells) of the protein.
CC See also Q20695-97.
SQ Sequence 1542 BP; 488 A; 277 C; 360 G; 417 T;

DB 3; Score 24; Match 71.4%; QryMatch 1.6%; Pred. No. 5.18e-01;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 149 cctggagctgcgaacatcgccgcactttgcaacaaggcagcagtagtca 204
    ||||| | ||| | ||||| | ||||| | ||||| | ||| | |||
Qy 980 CCTGGAGCCTCAGATCTGCCCGCGACTTTCAACAGAGCTCCCTGCATTCA 1035

RESULT 11
ID Q11195 standard; DNA; 36 BP.
AC Q11195;
DT DT 05-JUN-1991 (first entry)
DE Ballast Constituent coding sequence #2.
KW Ballast constituent; fusion protein; oligonucleotide library; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT repeat unit 7..9
FT /*tag= a
FT /notes= "can be present 4 to 8 times"
PN W09103550-A.
PD 21-MAR-1991.

```

[illegible][illegible]

```

FT repeat_region 7..30
* FT /*tag= b
* FT /note= "(DCD)8"
PN US5227293-A..
* PD 13-JUL-1993.
* PD 29-AUG-1989; 399874.
PR 29-AUG-1989; US-399874.
PR 23-APR-1992; US-838221.
PA (GENO ) GEN HOSPITAL CORP.
PA (FARH ) HOECHST AG.
PI Habermann P, Seed B, Stengelin S, Uhlmann E, Ulmer W;
DR WPI; 93-235119/29.
PT Fusion proteins for prodn. of e.g. pro-insulin - comprise gene
PT for desired protein and oligo-nucleotide(s) encoding ballast
PT protein
PS Claim 4; Column 29; 22pp; English.
CC This preferred mixed oligonucleotide encodes a ballast constituent
CC and is inserted between a regulatory region and the structural gene
CC encoding a desired protein. The short ballast component improves
CC protease resistance of the fusion protein while still allowing the
CC desired protein to adopt its correct conformation prior to cleavage
CC of the ballast constituent. Proteins which can be produced as fusion
CC proteins with a ballast region include pro-insulin and HMG CoA
CC reductase.
SQ Sequence 33 BP; 1 A; 10 C; 3 G; 1 T;

```

[illegible]

RESULT	15
--------	----

ID	Q51787	standard; DNA; 39 BP.
AC	Q51787;	
DT	20-DEC-1993	(first entry)
DE	Mixed oligonucleotide f19	encodes ballast constituent.
KW	Fusion protein; Ballast constituent;	pro-insulin production;
KW	recombinant protein production; HMG	CoA reductase;
KW	human 3-hydroxy-3-methylglutaryl-coenzyme	A-reductase;
KW	mixed oligonucleotide; ss.	
OS	Synthetic.	
FH	Key	Location/Qualifiers
FT	repeat_unit	4..6
FT	/*tag= a	
FT	/rpt_type= tandem	
FT	/note="can be repeated y times, where y is 4-11,	
FT	provided that y+z is 6-12"	
FT	repeat_region	4..36
FT	/*tag= b	
FT	/note="(DCD)11"	
FT	repeat_unit	36..38
FT	/*tag= c	
FT	/rpt_type= tandem	
FT	/note="can be repeated z times, where z is 1-4,	
FT	provided that y+z is 6-12; N stands for	
FT	identical or different nucleotides,	
FT	excluding stop codons"	
PN	U85227293-A.	
PD	13-JUL-1993.	
PR	29-AUG-1989; 399874.	
PR	29-AUG-1989; US-399874.	

```
PR   23-APR-1992; US-838221.
```

PA (GEHO) GEN HOSPITAL CORP.
PA (FARH) HOECHST AG.

PI Habermann P., Seed B, Stengelin S, Uhlmann E, Ulmer W;
PT WPt ; 93-235119/29.

DR Fusion proteins for prodn. of e.g. pro-insulin - comprise gene
PT for desired protein and oligo-nucleotide(s) encoding ballast
protein

PS Claim 9; Column 30; 22pp; English.

CC This preferred mixed oligonucleotide encodes a ballast constituent
CC and is inserted between a regulatory region and the structural gene
CC encoding a desired protein, esp. pro-insulin. The short ballast
CC component improves protease resistance of the fusion protein while
CC still allowing the desired protein to adopt its correct conformation
CC prior to cleavage of the ballast constituent.

SQ Sequence 39 BP; 1 A; 11 C; 1 G; 1 T;

DB 7; Score 22; Match 25.0%; QryMatch 1.5%; Pred. No. 4.35e+00;
Matches 7; Conservative 18; Mismatches 3; Indels 0; Gaps 0;

Dbb 6 ddddddcccccccddddcccdccccd 33
::l:l: ::ll: :ll::l:l:

Cp 1336 AACATCTGGAGCAGATCGACGTGCTT 1309

Search completed: Thu Sep 28 00:05:27 1995
Job time : 97 secs.

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MParch nn n.a. - n.a. database search, using Smith-Waterman algorithm

```
Run on:      Wed Sep 27 23:49:23 1995;   MasPar time 782.94 Seconds
           1107.806 Million cell updates/sec

Tabular output not generated.
```

```

>US-08-252-491-1
Description: (1:1486) from US08252491.seq
Perfect Score: 1486
N.A. Sequence: 1 CCTCGTGGCGGTCTCGGC.....TATAAATTTGAAATCACTA 1486
Comp: GGACAGCGCCAGACTCCG.....ATATTTAACTTTTAGTGT

```

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 267821 seqs, 291840210 bases x 2

Database: embl-new6

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
BCT	EST1	EST2	EST3	EST4	EST5	FUN	INV1	INV2	MAM	ORG	PLN	PRI	PRO	ROD	STS	SYN	UNC	VRT	VIR

Database: genbank89

Database:

genbank-new6

76	BCT
77	EST1
78	EST2
79	EST3
80	EST4
81	EST5
82	EST6
83	EST7
84	EST8
85	INV
86	MAM
87	PHG

88	PLN
89	PR1
90	ROD
91	STS
92	STR
93	SYN
94	UNA
95	VRL
96	VRT

Database: u-emb143.89
97 ALL

Statistics: Mean 11.614; Variance 3.738; scale 3.107

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1486	100.0	1486	57	MUSTHROA	0.00e+00
2	676	45.5	1795	50	HUMMLCPL	0.00e+00
3	616	41.5	1341	47	HSU11025	0.00e+00
4	602	40.5	1062	51	HUMTHROMB	0.00e+00
5	328	22.1	7666	51	HUMTA	0.00e+00
6	328	22.1	6163	51	HUMTHROWA	0.00e+00
7	26	1.7	3379	41	SCILVS	4.55e-03
8	25	1.7	335	44	HSDNASDAZ	2.64e-02
9	24	1.6	1525	59	S52337S1	1.46e-01
10	24	1.6	134226	71	IHICG	1.46e-01
11	24	1.6	3940	31	TRECRE1	1.46e-01
12	24	1.6	1514	57	MUSOSF1	1.46e-01
13	23	1.5	53222	50	HUMPKD1GEN	7.59e-01
14	23	1.5	329	67	ADRHUMJ	7.59e-01
15	23	1.5	5322	73	CHKOVALSA	7.59e-01
16	23	1.5	53222	89	HUMPKD1GEN	7.59e-01
17	23	1.5	675	75	S82572S1	7.59e-01
18	23	1.5	350	44	HSDINURPA	7.59e-01
19	23	1.5	875	62	G05108	7.59e-01
20	23	1.5	53222	13	HSPRD1GEN	7.59e-01
21	23	1.5	38139	85	CELF45E1	7.59e-01
22	23	1.5	38139	8	CELF45E1	7.59e-01
23	23	1.5	84539	44	HSAB1GR3	7.59e-01
24	22	1.5	1468	48	HUMC2PHTYR	7.59e-01
25	22	1.5	426	6	HS971139	3.71e+00
26	22	1.5	5174	44	HSCOL4A3	3.71e+00
27	22	1.5	2866	37	ANUAPA	3.71e+00
28	22	1.5	1521	48	HUMC1PHTYR	3.71e+00
29	22	1.5	417	6	HS941138	3.71e+00
30	22	1.5	970	47	HUMAAFA	3.71e+00
31	22	1.5	2246	48	HUMCD203	3.71e+00
32	22	1.5	1260	65	SYNBCTCONS	3.71e+00
33	22	1.5	8647	52	S45332	3.71e+00
34	22	1.5	1475	52	S62885	3.71e+00
35	22	1.5	3505	32	BTU03420	3.71e+00
36	22	1.5	467	4	HS586134	3.71e+00
37	22	1.5	970	52	S62884	3.71e+00
38	22	1.5	467	78	R69586	3.71e+00
39	22	1.5	1475	47	HUMAPB	3.71e+00
40	22	1.5	885	55	MUSCSCCE	3.71e+00

41	22	1.5	300	56	MUSINTZ2	Mouse integration sit	3.71e+00
42	22	1.5	1722	52	S57235	CD68=110kda transmem	3.71e+00
43	22	1.5	600	64	HUMUT6643	Human STS UT6643.	3.71e+00
44	22	1.5	14977	45	HSFURIN	Human fur gene, exons	3.71e+00
45	22	1.5	417	80	R75941	yi62a02.r1 Homo sapie	3.71e+00

ALIGNMENTS

RESULT 1	MUSTHROA	1486 bp	mRNA	ROD	06-JUL-1994
LOCUS	Mus musculus thrombopoietin mRNA, complete cds.				
DEFINITION	L34169				
ACCESSION	thrombopoietin.				
KEYWORDS	Mus musculus cDNA to mRNA.				
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.				
REFERENCE	1 (bases 1 to 1486)				
AUTHORS	Lok, S., Kaushansky, K., Holly, R.D., Kuijper, J.L., Lofton-Day, C.E., Oort, P.J., Grant, F.J., Heipel, M.D., Burkhead, S.K., Kramer, J.M., Bell, L.A.N., Sprecher, C.A., Blumberg, H., Johnson, R., Prunkard, D., Ching, A.F.T., Mathewes, S.L., Bailey, M.C., Forstrom, J.M., Buddle, M.M., Osborne, S.G., Evans, S.J., Sheppard, P.O., Presnell, S.R., O'Hara, P.J., Hagen, F.S., Roth, G.J. and Foster, D.C.				
TITLE	Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo				
JOURNAL	Nature 369, 565-568 (1994)				
MEDLINE	94261207				
COMMENT	NCBI gi: 508540				
FEATURES	Location/Qualifiers				
source	1..1486				
CDs	/organism="Mus musculus" /sequenced_mol="cDNA to mRNA" 174..1244 /note="NCBI gi: 508541" /codon_start=1 /function="c-mpl receptor ligand" /evidence=experimental /product="thrombopoietin" /translation="MELTDLLLAAILAVARLTSSVPACDPRLINKLRDLSLLH SRLSQCPDVPDLSIPVLPVAVDSLGEMKTOEQSQADILGAVSLLEGVMAARGOL EPSCLSLLGQSGVRLLCALGCLLTQPLQGRTHAKDPNALFLSIQQLLRGV RFLLVGEPTLCVRLTPTTAVPSSTSQLTKLNKFPNRTSGLETFNSVTATAGPOL LSRLQGRVKITPGQINQTSRSPVQISGVYLNRTGPNVNGTGLFAGTSLTLEASDTS PGAFNKGSIAPNLQGLPPSPSLAPDQHTPPFPSPALPTTHGSPPLQHLFPDPSTTM PNSTAPHPVTWYHPHNLISQET"				
BASE COUNT	331 a	484 c	337 g	334 t	
ORIGIN					
DB 57;	Score	1486;	Match	100.0%;	QryMatch 100.0%; Pred. No. 0.00e+00;
Matches	1486;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
Db	1	cctcgtgccggtcctgagcccttctccaccggacagagctcttgcccaactctctcc	60		
Qy	1	CCTGTCGCCGCTCTGAGGCCCTTCTCCACCGGACAGAGTCTTGGCCGACCTCTCC	60		
Db	61	caccgcactctgccaaagacagaagctcaagccgctccatgccccccaggaaga	120		
Qy	61	CACCCGACTCTGCCAAAGACAGCAGAGCTCAAGCCGCTCCATGCCCCAGGAAGA	120		
Db	121	ttcagggagagggcccatcacaggaggacacttcagttagacacctggccagaatggagc	180		
Qy	121	TTcagggagagggcccatcacaggaggacacttcagttagacacctggccagaatggagc	180		

Db 181 tgaactgattgtctcctgagggccatgcttcttgacgtgcaagaactaactctgcacgccc 240
 Qy 181 TGACTGATTGTCTCTGCGGGCATGCTCTTCGAGTGGCAAGACTTAACACTCTGTCCAGCC 240
 Db 241 ccgtagctcctgctgtgacccagaactcctaaataaactgctgagtgactccacactcc 300
 Qy 241 CCGTAGCTCTCTGCTGTGACCCAGACTCTCTAAATAAATGCTGCGGTGACTCCACCTCC 300
 Db 301 ttacacagcagactgagtcagctgtcccaagctcgaccccttctcctcctgttctgctgc 360
 Qy 301 TTACAGCCGCACTGAGTCAGTGTCCAGACCTCGACCCCTTTGCTATCCCTGCTCTCTGCTGC 360
 Db 361 ctgctgtggactttagctcgtggagatggaaacccagacggaacagagcaagcaggg 420
 Qy 361 CTGCTGTGGACTTTAGCTGGGAGAAATGGAAACCCAGACGGAACAGAGCAAGGACAGG 420
 Db 421 acattctagggcagctgtccctctactgagggagtgatggcagcagagacagctgg 480
 Qy 421 ACATTCTAGGGGCACTGTCCCTTCTACTGAGGGAGTGTGCGCAGCAGCAGGACAGTTGG 480
 Db 481 aacctcctgctctcatccctcctggacagcttctgtggcaggttcgcctcctcttgg 540
 Qy 481 AACCTCTCTCCCTCTCATCCCTCCTGGGACAGCTTCTTGCGGAGGTTCGCCCTCTCTTGG 540
 Db 541 gggccctgagggcctcctaggaacccagctcctctacagggcaggacacagctcaca 600
 Qy 541 GGGCCCTGCAGGGGCTCTAGGAACCCAGCTTCTCTACAGGGCAGGACACAGCTCACA 600
 Db 601 aggaacccaatgccctctcttgagcttgcaacaactgcttcgggaaaggctgcctcc 660
 Qy 601 AGGACCCCAATGCCCTCTCTGCTGAGCTTGCAACAACCTCTTCGGGGAAGGCTGCTCC 660
 Db 661 tgcctctggtgaaggtcccaacctctgtgcagcaggaacctgccaacacacagctgtcc 720
 Qy 661 TGCTTCTGGTAGAAGGTCCACCCCTCTGTCTCAGAGGAGCCCTGCCAACACACAGCTGTCC 720
 Db 721 caagcagtaacttctcaactctcaactcaacaagtctccaaacaggactcttgattgt 780
 Qy 721 CAAGCAGTACTTCTCACTCTCACACTAACAAGTCTCCAAACAGACTTCTGCAATTGT 780
 Db 781 tggagcagaacttcagtcacagcagaactgctggcctggactctgagcaggttc 840
 Qy 781 TGGACAGAACTTCAAGTGTACAGCCAGAACTGCTGGCCCTGAGCTTCTGAGCAGGCTTC 840
 Db 841 agggattcagagtcaagattactcctggtcagctaaatcaaacctccaggtcccaagtc 900
 Qy 841 AGGGATTCAAGATTACTCTCTGCTCAGCTAAATCAAACTCCAGGTCGCCAGTCC 900
 Db 901 aaatccttgatacctgaacagacacacagacctgtgaatgaaactcatgggctctttg 960
 Qy 901 AAATCTCTGGATACCTGAACAGACACAGGACCTCTGAATGAACTCATGGGCTTTTG 960
 Db 961 ctggaacctcaactcagacctggaagcctcagacatctcgcccgagctttcaacaag 1020
 Qy 961 CTGGAACTCACTTCAGACCTGGAGGCTCAGACATCTGCGCCGAGCTTCAACAAG 1020
 Db 1021 gctccctggcattcaacctccaggttgacttctcctctctccaaaccttgcctcctgatg 1080
 Qy 1021 GCTCCTGGCACTCAACCTCAGGGTGGACTTCTCTCTTCTCAAAAGCCTTGCTCTGATG 1080
 Db 1081 gacacacaccccttccctcctcaactgccttgccacacacacacatggatctccacccagc 1140
 Qy 1081 CACACACACCCCTTCCCTCTACCTGCTGTCGCCACACCCCATGATGATCCACCCAGC 1140

Db 1141 tcaacccctgttctcctgacccctccaccacatgcctaaactctaccgccctcatccag 1200
 Qy 1141 TCCACCCCTGTTCCTGACCTTCACACCATGCTTAACCTTACCGCCCTCATCCAG 1200
 Db 1201 tcaaatgtaccctcatccaggaattgtctcaggaataatagcggcgacactgccc 1260
 Qy 1201 TCAAAATGTACCCCTTATCCAGGAATTTGTCTCAGGAACATAGCGGGGACTGGCCCA 1260
 Db 1261 gtgagcgtctgcagcttctcctggggacacagcttcccccaggagcgtgagagcagctgc 1320
 Qy 1261 GTGAGCGCTGTGAGCTTCTCTGGGGGACAAGCTTCCCCAGGAAGGCTGAGAGGCACTGC 1320
 Db 1321 atcgtctccagatgtctcgtcttccacaaaggccctggggaaggatcacacagcactg 1380
 Qy 1321 ATCTGCTCCAGATGTCTGCTTTCACCTAAAGGCCCTGGGGAAGGATACACAGCACTG 1380
 Db 1381 gagattgaaaaatttagagactattttttaacctatcagcaatattcatcagagca 1440
 Qy 1381 GAGATTGTAATAATTTAGGAGCTATTTTTTTTAACCTATCAGCAATATTTCATCAGACCA 1440
 Db 1441 gctagcagcttctgctctatttcggtataaattgaaaatcacta 1486
 Qy 1441 GCTAGCGACTTTGGTCTATTTGGGTATATAATTTGAATCACTA 1486

RESULT 2
 LOCUS HUM1.CMPL 1795 bp mRNA PRI 07-JAN-1995
 DEFINITION Human c-mpl ligand (ML) mRNA, complete cds.
 ACCESSION L33410
 KEYWORDS c-mpl ligand.
 SOURCE Homo sapiens cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
 Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
 REFERENCE 1 (bases 1 to 1795)
 AUTHORS de Sauvage, F.J., Hass, P.E., Spencer, S.D., Malloy, B.E., Gurney, A.L.,
 Spencer, S.A., Darbonne, W.C., Henzel, W.J., Wong, S.C., Kuang, W.-J.,
 Oles, K.J., Hultgren, B., Solberg, L.A.Jr., Goeddel, D.V. and
 Eaton, D.L.
 TITLE Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl
 ligand [see comment]
 JOURNAL Nature 369 (6481), 533-538 (1994)
 MEDLINE 94261202
 COMMENT NCBI gi: 506826
 FEATURES
 source
 location 1..1795
 /organism="Homo sapiens"
 /sequenced_mol="cDNA to mRNA"
 216..278
 /gene="ML"
 /codon_start=1
 216..1277
 /gene="ML"
 /note="erythropoietin homology domain bp 66..522; NCBI
 gi: 506827"
 /codon_start=1
 /product="c-mpl ligand"
 /translation="MELTELLLVVLLLTARLTLSPPACDLRVLSKLRLDSVHLH
 SRISQCEVHPITPVLPVDFVDFSLGEMKQMEETRAQDILGAVTLLEGVNAARGL
 GPTLSSLLGSLGQVRLILGALSLGTLPPQGRTHAKDPNAIFLSFQHLIRKGV
 RFMLVGGSTLCVRRAPPTTAVTSRTSLVLTINELPNRTSLGLETNFTASRTTGSCL
 LKWOQGRKIPGLINSLDQIPYLNRIHELINGTRGLFFGPSRTIGADPISL
 GTSDTGSLPPNLPAGVSPSPHTPTQVTLFFLPPTLPVTVVQLHPLLPDPSAPTTP


```

mat_peptide      TSPLLNTSYTHSQNLSEOG*
                  279..1274
                  /gene="ML"
                  /codon_start=1
                  /product="c-mpl ligand"
BASE COUNT      420 a 586 c 382 g 407 t
ORIGIN
DB 50; Score      676; Match 80.7%; QryMatch 45.5%; Pred. No. 0.00e+00;
Matches 967; Conservative 0; Mismatches 219; Indels 12; Gaps 7;
Db 85 cettggccgcctttggccaccactactctgccagaagtgc-aagacct-aagccgcc 142
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 42 CTTGGCCGACCTCTCTCCACCCCACTCTGCCGAAAGACAGACAGACTCAGCGCCG 101
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 tccatggccagggaagattcaggggagagcccaaacagaggagccagccagccaga 202
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 102 TCCATGGCCCCAGGAAGATTACGGGGAGAGGCCCCCATACAGGGAGCCACTTCAGTTAGA 161
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 caccggccagaaatggagctgaatgctcctctgtgtatgctgcttctcctaactgc 262
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 162 CACCT-GGCCAGAAATGAGAGTGAATTTGCTCTCTGCGGCGCATGCTTCTTCGATGGC 220
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 aaggctaacctgtccagccggctcctcctgtgtgacctccagctcagtaaac 322
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 221 AAGACTAACTCTGTCCAGCCCGGTAGCTCTGCTGTGACCCCGACACTCTTAATAAACT 280
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 gcttgactcctccttccacagacactgagccagtgccagaggttcacccctt 382
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 281 GCTGGTGACTCCGACCTCTTCACAGCCGACTGAGTCACTGTCCGAGCTGCGACCTTT 340
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 gctacacctctctgctgctgtgactttagcttgggagaatggaaaaccagat 442
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 341 GCTATCTCTGTTGCTGCTGCTGTGGAATTTAGCCCTGGGAGATGGAAACCCAGAC 400
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 ggaggagaccaggcacagacattctggagcagtgacacctctgctggaggagtgat 502
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 401 GGAACAGAGCAAGGACAGGACATTTAGGGGCACTGTCCCTCTTACTGGAGGAGTGAT 460
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 ggcagcaggggacaactgggaaccacttgcctctcatccctcctggggcagctttcgg 562
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 461 GGCAGCAGGAGGACATTTGGAAACCTCTGCTCTCATCCCTCTGGGACAGCTTTCTGG 520
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 acagtcctgtctccttggggccctgcagagcctccttggaaaccagcttctccaca 622
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 521 GCAGCTTGCCCTCTCTTGGGGGCCCTCGAGGGCCCTCTAGGACCCAGCTTCTCTACA 580
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 gggcaggaccacagctcacagagatcccaatgccatctcctgaggttccaaacctgct 682
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 581 GGGCAGGACACAGCTCACAGGACCCCAATGCCCTCTTCTTCTGAGCTTGCAACAAGTCT 640
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 683 ccgaggaagagtgcttctctgactgttagagaggtccacacctctgctcaggcgggc 742
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 641 TCGGGAAAGGTGCGCTTCTCTGTTCTGTTAGAGGTCCCAACCTCTGCTGTACAGCGGAC 700
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 743 cccaccacacagctgtccccagcagaacctctctagctcctcacactgaacagctccc 802
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 701 CCTGCCAACACAGCTGTGCCAAGCAGTACTTCTCACTCTCTCACACTAAACAAGTTCCC 760
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 803 aaacaggactcttgattgttgagacaaacttcactgctcctcagccagcagaactggctc 862
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 761 AAACAGGACTTCTGATTGTTGGACGCACTTCAGTGTACAGGCCACAGCTGCTGGCCC 820
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 863 tgggctcttgagtgaggcaggggattcagagccaagatt---cctggtctctgtaacca 919

```

```

Qy 821 TGGACTTCTGACGACGCTTCAGGGAATTCAGAGTTCAAGATTACTCTGCTCAGCTAAATCA 880
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 920 aactccaggtccctggaccaaatccccgggataacctgaacagagatacaagaactcttga 979
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 881 AACCTCCAGGTCCCGCAGTCCAAATCTCTGGATACCTGAACAGGACACACGACCTGTGAA 940
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 980 tgaactcgtggactcttcttggacctcagcaggaccttaggagcccgagacattc 1039
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 941 TGGAACTCATGGGCTCTTCTGGAACTTCACTTCAACCCCTGGAGCCTCAGACATCTC 1000
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1040 etcaggaaatcagacacagctccttgcacccaacctccagcctggatattctcttc 1099
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1001 GCCGGAGGCTTCAACAAAGGCTCCCTGGCATTCACCTGCGAGGTGGACTTCTCTCTC 1060
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1100 ccaacaccatctcctcaggacagt-a---tagctcttccctcttcca---ccacactt 1153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1061 TCCAAGCTTGTCTCTCATGATGACACACACCCCTTCCCTCTTCACTGCTTGCACCCAC 1120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1154 gccacccctctggtccagctccacccctctgcttctgaccttctgctccaagccac 1213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1121 CCATGATCTCCACCCAGCTCCACCCCTGTTCTGACCTTCCACCAACCATGCTTAA 1180
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1214 cctaacagcctctttaaacaatcctacaccactcccagaatctgtctcaggaa 1271
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1181 CTCTACGGCCCTCATCTCAGTCAATGTACCCCTCATCCAGGAATTTGTCTCAGGAA 1238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT          3
LOCUS           HSU11025      1341 bp      mRNA              PRI      15-OCT-1994
DEFINITION      Human megakaryocyte growth and development factor (MGDF) mRNA,
complete cds.
ACCESSION       U11025
KEYWORDS        .
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryotes; Hyperchondria; Eukaryote crown group; Metazoa/Eumycota
group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia;
Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopherygii;
Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 1341)
AUTHORS         Bartley,T.D., Bogenberger,J., Hunt,P., Li,Y.S., Lu,H.S., Martin,F.,
Chang,M.S., Samal,B., Nichol,J.L., Swift,S. et.al.
TITLE           Identification and cloning of a megakaryocyte growth and
development factor that is a ligand for the cytokine receptor Mpl
JOURNAL         Cell 77 (7), 1117-1124 (1994)
MEDLINE         94291201
REFERENCE       2 (bases 1 to 1341)
AUTHORS         Samal,B.B.
TITLE           Direct Submission
JOURNAL         Submitted (18-JUN-1994) Babru B. Samal, Developmental Biology,
Angen Inc., Angen-center, Thousand Oaks, CA 91320, USA
COMMENT         NCBI gi: 511223
FEATURES        Location/Qualifiers
source          1..1341
                /clone="HFL-IVC7"
                /organism="Homo sapiens"
                /tissue_type="liver"
                /dev_stage="fetal"
                36..1097
                /gene="MGDF"
                /note="NCBI gi: 558078"
                /codon_start=1
CDS

```


<hr/>							
REFERENCE	Catarrhini; Homnidae; Homo.						
AUTHORS	1 (bases 1 to 335)						
	Zabarovsky,E., Kashuba,V., Pettersson,B., Petrov,N., Zakhar'yev,V.,						
	Gizatullin,R., Lebedeva,T., Bannikov,V., Eriandson,R., Uhlen,M.,						
	Kiselev,L. and Klein,G.						
TITLE	Shot-gun sequencing strategy for long range genome mapping: first results						
JOURNAL	Unpublished						
REFERENCE	2 (bases 1 to 335)						
AUTHORS	Zabarovsky,E.						
TITLE	Direct Submission						
JOURNAL	Submitted (01-APR-1993) to the EMBL/GenBank/DDBJ databases.						
	Zabarovsky E., Karolinska Institute, Tumor Biology, S-104 01 Stockholm, Sweden						
COMMENT	NCBI gi: 608819						
FEATURES	Location/Qualifiers						
source	1..335						
	/organism="Homo sapiens"						
	/clone="NL122060"						
	/cell_line="MCH903.1"						
	/clone_lib="NL1"						
	/chromosome="chromosome 3"						
	/germline						
BASE COUNT	38 a 149 c 99 g 16 t 33 others						
ORIGIN							
DB 44:	Score 25; Match 39.3%; QryMatch 1.7%; Pred. No. 2.64e-02;						
Matches	22; Conservative 17; Mismatches 17; Indels 0; Gaps 0;						
Db	279 agggddgdnckcyvvvvccccaanacckmmctttcccvmvggdtdtytdg 334						
	:::: : :: : : ::: :: :						
Cp	685 AGGTGGGACCTTCACAGAGCAGAGCGCACCTTTCCCGACAGCATTTGTG 630						
<hr/>							
RESULT	9						
LOCUS	S52337S1 1525 bp DNA						28-MAY-1993
DEFINITION	OSF-l=pleiotrophin [mice, DBA/2J, liver, Genomic, 1525 nt, segment 1 of 5].						
ACCESSION	S52337						
KEYWORDS	.						
SEGMENT	1 of 5						
SOURCE	mice liver DBA/2J.						
ORGANISM	Mus sp.						
	Unclassified.						
REFERENCE	1 (bases 1 to 1525)						
AUTHORS	Kato,K., Takeshita,S., Sato,M., Ito,T. and Amann,E.						
TITLE	Genomic organization of the mouse OSF-1 gene						
JOURNAL	DNA Cell Biol. 11 (10), 735-743 (1992)						
MEDLINE	93090271						
REMARK	GenBank staff at the National Library of Medicine created this entry [NCBI gibseq 120279] from the original journal article. This sequence comes from Fig. 4. Map location: 7q22-qter. NCBI gi: 298209						
COMMENT	Location/Qualifiers						
FEATURES	1..1525						
source	/organism="Mus sp."						
	/note="mice"						
BASE COUNT	402 a 350 c 395 g 378 t						
ORIGIN							
DB 59:	Score 24; Match 71.4%; QryMatch 1.6%; Pred. No. 1.46e-01;						
Matches	40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;						

Db	1388	cctcgagctctgcaagaacctgcgccgaactttgcacaaaggcagccagtcatca	1443
Qy	980	CCTGGAGCCTCAGACATCTCGCGGAGCTTTCAACAAGAGGCCTCCGTGCCATTCA	1035

RESULT	10		
LOCUS	IHICG	134226 bp	DNA VRL 15-JAN-1992
DEFINITION	Ictalurid herpesvirus 1 (channel catfish virus [CVI]), strain auburn 1, complete genome.		
ACCESSION	M75136		
KEYWORDS	DNA polymerase; SFI helicase; Zn-binding protein; dUTPase; glycoprotein; membrane protein; protein kinase; serine protease; terminase; thymidine kinase.		
SOURCE	Ictalurid herpesvirus 1 (strain auburn 1) (library: ATCC VR-665) DNA.		
ORGANISM	Ictalurid herpesvirus 1		
REFERENCE	Viridae; de-DNA enveloped viruses; Herpesviridae.		
AUTHORS	Harrison,P.T., Thompson,R. and Davison,A.J.		
TITLE	Evolution of herpesvirus thymidine kinases from cellular deoxycytidine kinase		
JOURNAL MEDLINE	J. Gen. Virol. 72, 2583-2586 (1991)		
REFERENCE	2 (sites)		
AUTHORS	Jenson,H.B.		
TITLE	GenBank Curator Program		
JOURNAL	Unpublished (1992)		
REFERENCE	3 (bases 1 to 134226)		
AUTHORS	Davison,A.J.		
TITLE	Channel catfish virus: A new type of herpesvirus		
JOURNAL MEDLINE	Virology 186, 9-14 (1992)		
COMMENT	NCBI gi: 331209		
FEATURES	Location/Qualifiers		
source	1..134226 /organism="Ictalurid herpesvirus 1"		
	/strain="auburn 1"		
	/sequenced_mol="DNA"		
	/tissue_lib="ATCC VR-665"		
repeat_unit	63..273 complement(467..472)		
polyA_signal	/note="ORF 1"		
polyA_signal	complement(583..588)		
repeat_unit	/note="ORF 1"		
CD5	609..748 complement(742..3330)		
	/note="ORF 1; NCBI gi: 331211"		
	/codon_start=1		
	/number=1		
	/translation="MDGLKEITAAVASLGGTDLSTYWNFDLGDMDSAGVVDV HPPEVGVSADMDVAESEDDEEGPTGEDESEADEDPEDVEDIPAAPVSP LPTSTTRGFGDADLGLGNLTADLAGNGLKGYAATLFESKTLPTLAADVEOKGALD PSVAKSVITLLSELNGLIGFSRLPTRQNKNAAWILMKVLGRALTLP.LTKTDGEV LSAINLVFGACKRVIVTSHTSDITKF.IIFENDDGTPFLGVDTVPALMKYIDPRKLKT ILNLKUNLGNVTVEWRKSCKTLWMREIRPVCFRFPVPTVESDNIIYATLDVKFR LLKGVSTQSSSYYVYPKSI.GDDFLNTKNTSLTLGDAP.LTVKLSFRKP.NTASQSGFK RMLTCLNLRMGRIELGPKSTLRSHEESLEEENVTCTCRMSLTRLASRAMSKP NTLIKNDVRMP.TSILENP.DFVRHGNTLP.IWILLFAASLRID.IVAKEWLSTEGDPR SIDYAKLTHQYLEIRLLSDAGIAKKGLSKULTATNNWLI.TEMSF.IVMYKGS.YCMPT VLST.IIDSNVNP.IDP.FSWGEIP.LMSTSGI.GRVEKVIVFTNSISVSRMLLSIDVEXAK WAIHAJKRAEOKA.EKESDAADKAP.KRRKEANGETPKKKOKEEKSEXPXKTK KAESCTACTAVTKKETKTP.QTSTNKBNKVS.SAEAGTA.VTP.DEANGKKSQVWK RKESAQHTTEKTKTAAKKVITPD.PSGHRP.SSKREYSQFEVSDDSDSEEVISKRP VTTTNMKSTTKKTLASNNDDDDGDDDFMTCAQGITGGGSGFLTSTAQRNRKSNQKQ		

```

CDS
3502..4725
/note="ORF 2; NCBI gi: 331212"
/codon_start=1
/number=2
/translation="MTGRHSPIHQTSPPRPVEKMPRFQREHGASHRYTPMGKRPAP
ANTGSHRGALLGRNGDPERKGEETAVSKLLKSGYIPETARTALIENVEKEGARGGE
EVMGCSLLPQHDSGRAIDDFRDMRSVRGEVASPIDELLFRVGNLLEVVEITTG
RPVPAGKLGAVVQVVTGIPKMIHVCLNMFELFGGRVTRGVSPGERRVTHTRCLGDS
TTPHCFWRTTHSDAQASDFGLNMIFELFGGRVTRGVSPGERRVTHTRCLGDS
FNALFAVHDPVTSRAFEARVTVQTSYPAWTFGDMWTPAPNNVGVWVTPGSMC
YDIRVDLNDPNKMSVTPAAAPPVGPKEPHEELEDAPWKPSTTIGSPAPNLCALY
TSGDVELHAPLLPL"
4886..5794
/note="ORF 3; NCBI gi: 331213"
/codon_start=1
/number=3
/translation="MAFSTSEKRIAEILGMSPATLSVRVGVGLPHDP1PWDSDEG
DDTMTGSGSERALGVETVARPHVYIGPVGICAVDDSDSDSPSPVGTGAERAGG
LPQAPILVVRVSGNERVVSERIEEIGRELARICELQWDEVDVAGAPVCTVGN
PWSVPGSLPPKSGSTTWNPPSPSDITCEIPSLPSSLVSPVPAPWISPVSD
APLNTSPNMITPTPGLRLDTPPTPLADLDGWANDWLYAPQLGQQLRIESGSS
PKRPLDDPDEVPSKRGPSRRLLN"
5668..5956
repeat_unit
repeat_unit
polyA_signal
6084..6241
6118..6123
/note="ORF's 2 and 3"
6158..6163
/note="ORF's 2 and 3"
6194..6199
/note="ORF's 2 and 3"
6234..6239
/note="ORF's 2 and 3"
6267..6272
/note="ORF's 2 and 3"
6513..7286
repeat_unit
CDS
7960..8499
/note="ORF 4; NCBI gi: 331214"
/codon_start=1
/number=4
/translation="MASIDVGETIGIQVEPWKRFVCHGCFPPSEVTARGAGSVSL
RVVHYVSPERDEPSPRGVDEVDGACSEPTPRPEPRFRAIEMGKLVVLVSVCP
LRPALQRWVGAPARRHRSDSVARRAFEPWGRASRPTVLTPASGSDDDVDTFTGC
CGGPGHPVDVCHCDQDWL"
8557..8562
/note="ORF 4"
8642..8647
/note="ORF 4"
8785..9471
/note="ORF 5; NCBI gi: 331215"
/codon_start=1
/number=5
/product="thymidine kinase"
/translation="MALREGTPARELPSELVFCVEGNIGCGKSTLWKALMERVAGSG
VVNVEFVDWNNHGNKYLELSYDPTGVAVPFQNLVDSVNVQRIONPDIWERSP
MSATRVFCVNGSGVTPATLPQMAARGEAMRTIATRPVYLELPFEECLRMRR
RDRTGAGVGDYLRILHERVEAWLSSAEDVERVDASRSREETIDRVIELCRHRPL
RAPLTRKSOL"
9535..9951
/note="ORF 6; potential membrane protein; putative; NCBI
gi: 331216"
/codon_start=1
/number=6
/product="membrane protein"

```

```

/translation="MNSLTIFLLSGIATYHAYVADGTGSSSVTAGDSGVVLMVIG
ALLTLMTIPITIGLFGIVYRTRASIEEMRGILMQIHLRLITCDQGRNRGDVELGAGS
LLTISOPPSSYAEALLMEPEVQEQGVPLEAEIRV"
10043..10048
/note="ORF's 5 and 6"
10248..10267
repeat_unit
10298..10326
repeat_unit
10420..10614
10863..11285
/notes="ORF 7; potential membrane protein; putative; NCBI
gi: 331217"
/codon_start=1
/number=7
/product="membrane protein"
/translation="MAAVILERAEEVAPGEARVGVPIIAEYVYRALTSDHEMRAFYET
CAVSFALEMLIIVLHASRHEGSTTRCTDAHTQTEGSTTRCTDAHTQTEGSRDQGS
MTEPADDLTRPPLGHGRQIPVLRRLVLDRLRIDYSL"
11059..11152
11499..11504
/note="ORF 7"
11658..11751
11767..11898
12462..13160
/notes="ORF 8; potential membrane protein; putative; NCBI
gi: 331218"
/codon_start=1
/number=8
/product="membrane protein"
/translation="MKCPARVDIFKREIKTRELEPTKKNVYHGLTGPISLPPHPV
IRQENWNRLSAVIGFFLLVSTSGSAASIPSPAPDPTRESTPGEPRHRLST
ETPTPEPBGDGSPEVLHVVTGVPVRDRDPILERLAEILAEILSHQLLTPGTGR
EDEDEVFARALAAAEITAGSVADRVMMKATLSCLMLVTSILVFAGVALWVIVARHGFR
VIPHERSDHSPGL"
13197..13673
/notes="ORF 9; potential Zn-binding protein; related
sequence to ORF 10, ORF 11; putative; NCBI gi: 331219"
/codon_start=1
/number=9
/product="Zn-binding protein"
/translation="MATRPKVRPRDALVPPKVKRIKSVIFACGHRKRCGLGTRPLG
DSNGCPGCRARDVFAPLYGSGPLPPCPVCGRAVVGPTVREACGHVTCNACETEACA
VDRLCICGGGRLLVAICDPYPPGPHWRGPRTRPEAEAVQSRGSSSEDACTCAP"
13714..13733
13776..13781
/notes="ORF's 8 and 9"
14076..14531
/notes="ORF 10; potential membrane protein; related
sequence to ORF 9, ORF 11; putative; NCBI gi: 331220"
/codon_start=1
/number=10
/product="membrane protein"
/translation="MTWKGCLLAFQVTLIVIVGVVAVGVALSKGRAPGDDLVTLAL
NELAGDEVDVGRISDAGAAINTGLLRGLLDHLSRNSGSLVAADALYERVLRLRDI
FVNGLVAAFGNFWLSFIIMYTCARTVRSKKVAPAPIHSPPTMMSPYI"
14539..15384
/notes="ORF 11; potential Zn-binding protein; related
sequence to ORF 9, ORF 10; putative; NCBI gi: 331221"
/codon_start=1
/number=11
/product="Zn-binding protein"
/translation="MRCIRAAPADGPTYPIIMSPVFCFQCGEAVFPSPNRSCKHV
FCFFCTPCECGCGGKRLIENEYIALYATKPPFPAPMCRMTAGFWLMPNGCHWVE
PRVVLEDLTVAVITVVALVETRGAPGVRRARGKEGATLPLPLDDIVELGG
AIEAAGKAVVGGFIVRTLVELVRYHDTIAKTFPVMAPRFSGISALKEILLSRPIPG

```

CDS
 15391..16290
 /note="ORF 12; potential Zn-binding protein; putative;
 NCBI gi: 331222"
 /codon_start=1
 /number=12
 /product="Zn-binding protein"
 /translation="MTSPREITVRLSCGHPILRVPDGEFCFGDGAYCCDCDERAVFA
 PDHLTVTLKVAPEGLDAUTEKSNKLAKAIVTCGCGDVLVQPVHERCGHWCRCF
 RUKCAEGGLARTAVRPLDLERLQESLFGHTRGLEYSDKFTAEIWMGPQKVEVILLY
 DCAIYNQDLVVTGISAADRAEKPIDLVDIVLNPQSPVCTIPYRLFLLKANA
 EMGVQMSDLVTLGARNILVEYGAVPGLDRSLLEFLRYLVIPGFGAETFFALTEIF
 LSLIAECTWQDSRCGRDACS"

polyA_signal
 16329..16334
 /note="ORF's 10, 11 and 12"
 16354..16602
 /note="ORF 13; NCBI gi: 331223"
 /codon_start=1
 /number=13
 /translation="MAFLPFLNCONPMSLILCGGCDLISCCRGCGQMPARQPIY
 PYGSPMGAVHYPPVPAQPVVRGPRVVRQGERPVDLGR"

repeat_unit
 polyA_signal
 16665..16716
 /note="ORF 13"
 16740..16745
 /note="ORF 14"
 16744..16749
 /note="ORF 13"
 17047..17356
 complement(17076..17081)
 /note="ORF 14"
 complement(17128..17133)
 /note="ORF 14"
 complement(17179..17184)
 /note="ORF 14"
 complement(17231..17236)
 /note="ORF 14"
 complement(17298..17303)
 /note="ORF 14"
 complement(17346..18521)
 /note="ORF 14; related sequence to ORF 15, ORF 16; NCBI
 gi: 331224"
 /codon_start=1
 /number=14
 /product="protein kinase"
 /translation="MAEKLIPSCGOMFLPARVLREGFLRARPVAVDGGFKLTIGPLT
 VATRWVWVYVYHNEATDNLKGNPFPEEPCDVLDFPGVACVNEATFLIKKKYG
 SYLHFLPRLHTRGNIEFGSPLSAQMAVLKRTFGGQAPDRPGVPMYAWETA
 HFLDAAPKLEIMESGTRNRSKQAWYFSEGLYGSMEGYFRERATVDICATILAD
 LTGKLLALRKGIYHGDIAKENIIMRSRCPCKLIDFESHGPGETWTSFWYDRTFF
 WNP IGTGAATAPERSGRGRNAGVGVFIQGLIALINWERMVFNHTIKGGY
 RAHVIKVTKARTGLDRGGARTLARVDELIGIARCLERDPAMRP SLETLVDFSKI"
 18557..115670
 /note="unique region"
 complement(18768..18773)
 /note="ORF 15"
 complement(18773..18778)
 /note="ORF 15"
 complement(18837..18842)
 /note="ORF 15"
 complement(18854..18859)
 /note="ORF 15"
 complement(18866..20008)
 /note="ORF 15; related sequence to ORF 14, ORF 16; NCBI
 gi: 331225"

misc_feature
 polyA_signal
 polyA_signal
 polyA_signal
 polyA_signal
 CDS

/codon_start=1
 /number=15
 /product="protein kinase"
 /translation="MAAVNWKDEPYPEKPTRRHLSFGPARLPTGDWDMINTYKPE
 ARCMFLTNPNLWMSGEDVLGLLPAGVPTEKIFVKEVYPLKGFQMEIPVKVAGCL
 LFGFSPLSRTGMATLKATPVRGGYAPAGPIPMOVYAWEMACKIDAAPRLFKNELIV
 GEYFVRLATISECSAGDVGAYIRGGRPISIEAAAVKTRELASTLYLLAQNRYHGDV
 KIANVTYTPHGRGLIDFEMAHPLDMTMSGLGLEVP IKWDMVCTDEYRAPEGHGP
 FPDIISAAQLWQVGLMFLDVIIEI IENRNOGLWEQDPFGI AAREVVGIGRSLILAC
 EHRAFDLYITARGCLRTNPERPRILILLIAQLTKFIREVATQPEH"
 complement(20123..20128)
 /note="ORF 16"
 20139..20161
 complement(20171..21334)
 /note="ORF 16; related sequence to ORF 14, ORF 15; NCBI
 gi: 331226"
 /codon_start=1
 /number=16
 /product="protein kinase"
 /translation="MTSSCFIRAHPCVYVRCCFTRGCGIKVPCALITDFGLAWHAYY
 TDEALNWLADNPLMKSADVMDLLPGDMYVSEHVERKILSRVGPFEFDSFIPDIRV
 EECILFGSPLSSTGMAAKADILVYGYAPAPGIPMEVYAWEMAYKI GAAPRLLRW
 VLIEGEGIEQFATIAESGLRSLRDYICTHESGCKTPMDLGLVKNVHGLVAALGLML
 DSNLYHGDILKDLNLTVPFGPKYKLIIDFPAHFPEARMHPMEARGPITWLPMPGPAC
 NPPEHDEETPCRRRMKELGVYQJGLIMLETITVLDEALVRDENMEKRPDKRLVSA
 VMEENGSLAPHDRLLEGYLDLIGECILKRDTAVRMDLVITLMTLSLLIEKFKL"
 complement(22737..22742)
 /note="ORF 17"
 complement(22832..23728)
 /note="ORF 17; NCBI gi: 331227"
 /codon_start=1
 /number=17
 /translation="MIPPPIPVQLSFPICGSIRISISVPTGTHNALDRILSRSPKWK
 FTDYINGALACVGAAYAKALDNFGMRVSIVIGGVTVGLCAAFGPKKITELNALSQ
 MLAEATEDIKRERLTKALTSITLAASRAPINTWTRPRERFDALTCIPARMDDALCE
 MEFDTCGSGERGLHALHSLVIGELERLLVEGHVSTDLAOVYLTMLIVTLFRDDE
 FGKALLKVDLSIIALLESRLDVLLEEIQRVMSASVTRFHGVRPGFETEAFAMHDAIN
 RCGLYGTARGEPEVEEFGSTV"
 23762..24718
 /note="ORF 18; NCBI gi: 331228"
 /codon_start=1
 /number=18
 /translation="MKASQERSAARTASHVKEKKYMWASPRKRRDLAEKAAALWL
 EPSARLCLANPEPMVTLIHHYAAEVSTSRQYRGDFKGEPAFRFSPLDPTVNRFTIP
 PVSVLNPLAVSHTLLHGGIAMPKPCFVAFTENRSKFPAQAYELTVIFNEARVP IP
 QRTEWELATLSPEKTTVRIILNWSGTFEDQLEITERVLSRWADFSEYQLSRHSEF
 AVTSLDLRGYDSSGQYHAKMDSYDEVGVYVVRNLTASIPRRFTLVICHGTAVT
 ADPVLETDFTYKWLADDGDDGDDGDDGDDGDDGDDGDDGDDDE"
 24638..24712
 24748..24753
 /note="ORF 18"
 24842..25456
 /note="ORF 19; potential membrane protein; putative; NCBI
 gi: 331229"
 /codon_start=1
 /number=19
 /product="membrane protein"
 /translation="METMGPWKSIMFWIVLIVSGEGTDAVEIYMLEIHESAVPCVG
 SKNTMGCDLLAGGILGCGEMLHRDTPLRGGVINKYSRANPRFWPLSVMTILH
 VLIITACLTIIYIAKYRSRCRRPIERTKDPDPDERVRLYLDVRRRRYWP CSCTEGDD
 RTRLMPGDRVEYDGREKLVRSDDVTTHTLLTDPGDTGTIANE"
 25452..25457
 /note="ORF 19"
 complement(25562..25567)

polyA_signal
 repeat_unit
 CDS

polyA_signal
 repeat_unit
 polyA_signal
 CDS

polyA_signal
 repeat_unit
 polyA_signal
 CDS

CDS
/note="ORF 20"
complement (25576..25884)
/note="ORF 20; NCBI gi: 331230"
/codon_start=1
/number=20
/translation="MASSDRFPJFCCYRNCIFRVSVDVQPTIAHLRDFHGKPGCTSI
RFRCPCEIILRRPFLGETVDPENIRRAHVALIQNECRSPETQIMCIHFLLDPES
"
CDS
26089..26979
/note="ORF 21; NCBI gi: 331231"
/note="ORF 21"
/codon_start=1
/number=21
/translation="MSATGCPITEAANGCDDTGLCKTKNHDGAVPRTTCEPRSLICE
NRASSVITGLGPKPGVWDPRIHFHGVSESLPITDTPWGPPSPMDALGHGWTPS
GLKWCPPPPPTESRPSLELLAMHQKAQRDDTVTGETETAREDDVIFVEETNYN
NPDIIEVLKQVTECHGTIKRRYPMRVRRAPKRLVLDQVVDYFADSDDDTDAES
DDMSVLSDTCTDDASVSSCGSFITDGSSESESDASDETDDSDFDDELTSSES
EESESESESESESESE"
26926..26977
27099..27104
/note="ORF 21"
complement (27145..27150)
/note="ORF's 22 and 23"
complement (27146..31357)
/note="ORF 22; NCBI gi: 331232"
/codon_start=1
/number=22
/translation="MCP IRKPVNQWELSENKRNLCGTF GEMYVQVSTSPSOTAAAIV
KMFQYLRFTFOHLEITQASDAGDPVSRMDVLIQRIEELGTTVSDSDSVEFIEAHL
KYKIDFGNCSVTWIAOVKAADRAVTAATVGTGVMGAMGTLDRAMIDIKLILV
LPTTGTSPFVGVLPATPLVPYTAGSAGSPLDPTDRFGALFLGTLTPTATKK
VFQNDILNKVAQISVSVTAMNSFOAGLAKYKPRVPTADGALTARGIAPDRPQ
AKYETKVTATNMGFDVLIGCKVPCKSGDIESLVSGILLACKSLDELGALVGGISDG
AIVSRQMDQVRSILDSLVKTIYTDTRDRITAMTGCVHYTEANDLLRLVLDLIAKYG
EFOFTTISGLEPFWATIVSNTEGLPLVYAEAAVYAAERAVIIDEFKQMDHANNKVEV
EHI RERAAATOKOTEEHRAVDLOKAVETTAKMTDMKNGVEKTTADLRAQLSVANLSI
ESLKGTLKETTGLTTVMOLSKANQLSRVLAELQGLKGLQENEAULRDEQKQAQ
IANLRDLSAVTEALRHQKOTKLTETINTGVSTARVELEQRVELVAEKGS IREHR
MARMDPEAAKGP ROSEPRPSOTAVAHCALIRKNIHDPTNKRLLTNTNLEKSI
SKKELISSLITKGNALSQADAREQHAVIKREYDALRERGGTARTKKDALEAQS
KLITQESQALATEKLMVELEAAKRLCKSVTSKEVTEVRVNLITTEILARTKKTYD
ELRLNNAVLTAKORLOARVEALTDDNOSMKLLVNGGTLIDEANKRIITTLTORLATVS
GELVTATETLAPVSGSKRVNHQDELAKESATKSCENQTRALTRHATEMTAAG
ARDGLEALGRVRLTYDTETKLAGVERDVRVIRDELTRMQKASHIGADRLTHAEKV
AHDGMVNLAGLTARTTELRIQELVHEKRSRDAEQRHVASEAALREQIRTLTDERO
RLIATGVADLORENAELHAAITEAANTCADTKIQLENIADYNGLIAEYERLGRPP
GPVDPATTEP IAGYGQMDVTS GAARAP IDHDMFGISSLIGESQGGSLDHYGSY
AGRELSEESGDESVDVDPDSNGTRGDGCREDEDRQGETGESGEESESEES
EESDEDP SAGCPLELIGVPAGGIGADPEPARVPVSRPVTILLERGAHVAPKPTPR
VRGRPAARSHKRVHLISITRETLRLAALAGEPMSAEEIFQRINTDMETSSLSLNTV
TAADIPRQGLAVLRADTVAYETATPTVPSDRDQRHVELSKWAILIIMICAKHGIMNHE
EFTDGDLSWFKSYDNLSVTSQDSE"
complement (31281..32537)
/note="ORF 23; NCBI gi: 331233"
/codon_start=1
/number=23
/translation="MYIRINESRGESTLDRNLISASRVPVSVSSARFLSTCAESSRL
RVSTSLRSSRSLSFASVNSREITAAVPVTTIAGMEDVITITIGDITVYFLEWALISIN
DVLISLRKRGKFTTVDVTSNTSVTIAVPATMDPMERLFWTEGHHGVFMDLKGAT
YEMLCVLAAGVLSRMDPMP SLLDRDTSDELEIFVHLGRNTEINGATALWDRTVR
YAFVPSIELRFXKNWNAI LKGGDPRSVMRDVATVCTVGADALVLPVPGDGDGV
PVAVNRADRYLEELSERSTNTVTEILAAPTPFPPIPNFGICVLTRVKGGLCEVFI I
IKDSRMWEDRYTFAINTMIGNVISNGLIGITETTLAGLERTLOPPAPTSRPFVGVV
RDGSHTEAQPGVGLCKQAKPRNL"

CDS
32293..33294
/note="ORF 24; NCBI gi: 331234"
/codon_start=1
/number=24
/translation="MTSSIPAIVYVNGTAAISRELTI AKRERDELAREVETINLELSA
QVKNRALEETLGTDRDAISRLSKVSDPSLIRMYITKNOTLAQVRKTL EHLRLARE
TKGSRAGPATGTGYGPGPYRPVPIPADYPIHYSRVGPEYVLTWYATGCKQKN
ILDTYVRGVVTFEYFIVLSASAIFPVGMFDRQFVSSTGHDLDDYFCMKVALVATQ
YLTLINTSDSAQALAKTHNQRLQFTQFLRVQQQFQVLDVNIIRSAELGEVYTSAP
PLDEYTKRANLITAEAGDEDLKLILADDPILRVPAEKNKRRRAIGSSEELTACVE"
33342..34838
/note="ORF 25; NCBI gi: 331235"
/codon_start=1
/number=25
/product="SFI helicase"
/translation="MDVEIKYAGVQNEVWRVGVASIMMTHPGDGAIVLSGDAGCGKS
HAYQQLIEASRMGFAECVNITATTHKAVGVILQKQPCSTTYQSSIMASIDMQANFEF
KCLFDARHQAIEEMDRIVSNRTGLAAVRKKGHSCVTIANRCPVCVATIHESI RCTDRP
YPYFIARTWFIIDEYGIILSLSLKLILVALSKFLPGCYLIVFTGVSQLPSPHPQ
IMQPTLLRDMILINTYSLFINERVEDAAVAEAINLFQFNVITRPAVEMANARVIGANAV
DAEYCDVTRIFNGNHERDAYNTRHTWMEGRGAGRDVLTPIEHVGGSSASRCGALI
EYMCSTVKPIFNKATKGRVTYEGGVFVLPQTOERYTFFGLDDVNTAVVYGRDGRILY
ERQITKYGWTVSFFPIEPVAAINTYSAQGETLDRVIYAPPEKNYSMSSIRASAYAC
TRVTRTAIELSONSFSAKSVGRTNMPPEKLIHHRKSVEMGVYVRECPY"
34727..35146
/note="ORF 26; NCBI gi: 331236"
/codon_start=1
/number=26
/translation="MOLVREKRGAGHOHVPKRTTEPKVGRGDGICPERMSLLMMINPF
VDSVHLETVDALEAVYTGQPPFPDVRNDSRKYIMDRIHPWFVTCERPNTLITL
VSKIRDEAIVFGASIVYTTLTHLVPQREVNRKRIKII"
35148..36014
/note="ORF 27; NCBI gi: 331237"
/codon_start=1
/number=27
/translation="MASITGQLSPMYINNVTSAVQYIPVLEREDPRDLEVPKGVF
VP LSGAMTFPVYHGMSLFORPLGLVQTGGMLMSEPMVDVVP GKLYTVLNTGCDTLRV
GDHFVYQPPQEDTDAQASKPKGTIRICNNAASFNRMILPWGFMVAVKKIDP IVPRE
LMHQLTRDVTABSSSELLYGLAGYDILEKTSAAAGVFKDIAEKKA VTKEEVAAAT
ADQVDIFRKIVDEFSTMLNCVNFATIGMTHSSTLANPTPSLSQNVQWCMTCGTETPGL
QVHAFIKQFSGV"
35991..35996
/note="ORF's 24, 25, 26 and 27"
36035..36040
/note="ORF's 24, 25, 26 and 27"
complement (36075..36080)
/note="ORF's 28 and 29"
complement (36102..37874)
/note="ORF 28; NCBI gi: 331238"
/codon_start=1
/number=28
/translation="MSLPGRGTVKLTETRIIRVVRNGETGVYDTRAGSFETVSCQE
FEAAATVPVSVFPVCDRCFGTSLVEYPLTGFGTFVVGTCCTIFSPGDVDDRSIPAHMR
KYQQP IEAQTQMVQVAPGTIKYSHQIPMKGVLTGYHWHVMDRVYINNIIGTDESDTG
KRCVETFEADIPCALISGLTLDVGLNEVILECSVWVIPARRGCHAKLETRDTYSDGLE
KFCQSHATLPPTLLASFSTSSPERKTFYEAHVDA LNNYIKLLTIYSHKGETEIE
QYLIEGSKYSELIGEPSRVIDATMKAQIAEPQTHTGADQRQRPQDGP IYSPVDR
FPMTPYPAQFCGDLVSHYVPFPQSGGCGYPERVGVYPPQPPNPVYLP IYSG
QQYYPGHPQPHGHQQRSGGGLKAELETIGLAPKTNVQESLKSEISEILESELKQ
CGIKRAAGNIERNCDYDEEPRTKRARPEKTAFEAIVRAPYGFDFSTALTATTKIQVS
DTVEKLNKVIETLLTQSSAQAPLSTPQAAPVQPSLQPVPEP LAPQEPPEPPTGTSAP
TLEASLPQKRPVWSKAFETLNNI"
complement (37738..38499)
/note="ORF 29; NCBI gi: 331239"

```

/number=29
/translation="MTRKIDAPWLGRAEVLIPVHGHSHAVIEDVNAIAHRIENIHH
RSQSRMQLNGDIFDGVQLTILLGLEVDKCLSLVPGVGLLHAKKCGICGTRK
PDRLGFTYFGLPEEDHRSRHFQFNDFGYLLRSMDLDAYTAPITAPVPGVDCV
VIYKVVHPIPIPAKLPILALANVOYEMNTVQOOLIRTPKIRNSVVSROVTTRRKVR
EDRUSGEDIGAESKRGNGSVRYTGRELDSINSRI"
38368..38994
/notes="ORF 30; NCBI gi: 331240"
/number=30
/translation="MMDIFNSVSESINIINNSVAMPDREHEGAPSDKPGWNEFSGH
DFLWNSRSDIITSVQVDRTGSCPKASKNDOGTSLAEMESMWATQLODSAMSL
SWRPLDANWDVINYHESGICODPIEKLEYQIDTMDVHYRAGMAGDEFAWMTANG
SQFVSPNEKNICLSSLTTLGTLKLIQVRGDIRKMTTSKKNRGCTAALP"
39010..39015
/notes="ORF 30"
complement (39052..39057)
/notes="ORF 31"
complement (39053..39301)
/notes="ORF 31; NCBI gi: 331241"
/number=31
/translation="MYLQWFFPCPSFSFHGESVWVWVVDGRTGCRTHRHGRICIV
IQSAGNPVHATDYEAWLRMLRMGRIGRLVILYTSK"
39359..39763
/notes="ORF 32; NCBI gi: 331242"
/number=32
/translation="MKGATTDSDITSSVRALPAPVPGSGLELPLWLVHVSLLSLDQILM
GSKSDRPLRIGIGVSTVATVPDTPRMAHTEPVFLGTRSDLMFSPHNKNSSVLQ
PNDLSFLSFSVSPRNPDPNFASIAIVS"
39821..40693
/notes="ORF 33; NCBI gi: 331243"
/number=33
/translation="MEPISYVTITDYTAFAFWLLSTTTVDVDFWDRIGGFNLTASN
GRLLDLSRRAYVHNLRVHGHSFSELTGVGGVERTPLRALTLRAWNSEPERVLRITR
ELRTEKLHERGDYGRNLEVMCTNLAETPEIFETLCHRIIPELKRLVTTMKVYIT
AMYAGRDDULVQVTFYIRRSIEVDRELLEFIEGELKHSCLPISRKMTDKTFHVSOF
ITTTCTKELDRDDTARLVHRNCTGFETVVDRIWVGATMMVRRYTRAVESYDEFWYKY
CGWMDPTGCLPDE"
40536..41861
/notes="ORF 34; NCBI gi: 331244"
/number=34
/translation="MERVLRLQISTGSGRCGGIEERIERMSRTTNSGNTAGWIP
PPASFLMSTVLRVHSGSDGFFIHWKHVNTVNRQHVTRVMEHKTIGSNVVKRAT
AKFVDTKKRI SNVGSKSTRIEFTTREHSANYTANCKPLVQPTYKSYFNLLMSHGE
CTITHVDVNNIATYLYGVGNPMTSVWDSMKGRNLTPEFFSSINLAGIETNQ
LEFISMTINTQKLTHTTYLQGLKTLPTYSILLEVDTEKFMFNMIALFLECFINTPNKI
GMIWMNERYLENTEATSTETWNEYKYMLEKCAPLNGFSNAFAORTGQFYKN
CEWVHAPTFVAALAEAVLSYGSFLINTRKIKSFKELVWMLSVTTPDRLVTVQTDG
RTDYYRTGDRVITWMSHADFKVYASERLVLRKKINQYIKCKTGDHGDQD"
41707..42423
/notes="ORF 35; NCBI gi: 331245"
/number=35
/translation="MTAPMSIEPTGSGSPGLTELISKWRGLSDSFLEBRILTNISAQQ
EITAIKTENEARAKALROSLEDRAFWNKRIALCYKVAKRMAWVTEPGESTAPHL
EWHVSKGLGLGFLFISAGDOLLCPSAITPFIIDRGGPGTCLRLTFDOKLEGHLYTE
LISSGHGVSRGSDGNRAHFTGCTPVLPRKTLGKERRATLSAAIKQFKCLQVDHCT
LEGITESTAIKRAKILDLCE"
42425..42868

```

CDS

polyA_signal

polyA_signal

CDS

CDS

CDS

CDS

CDS

CDS

```

/notes="ORF 36; NCBI gi: 331246"
/number=36
/translation="MDGIFPSTDKMLYENVTIHRASGEYQTATLISHNHALPDSGS
PCDFEASIKKNGISLLTTRGPVTCIRKQIPEDKRVISLETIACLETIKVVGDCYI
TVTTGDVLEPPTPQAPPTPEAVKQPIATRRKSGLSALLMPF"
complement (42925..42930)
/notes="ORF 37"
42950..42995
/notes="ORF" s 32, 33, 34, 35 and 36"
complement (42990..42995)
/notes="ORF 37"
complement (43015..45027)
/notes="ORF 37; NCBI gi: 331247"
/number=37
/translation="MSGVKLGFSALISYTCVFNLRKKPIMMYRLGHQHQHPNT
DISIFRQKVPYPTDDAAINKLIMITRETYGIPKYITRLISGTFWAVIKRIYDNDRL
VFGNTTSCMMNGLDLVYADDCTKQLLEMPFHVDMQHLGAMSKSLGRTGTYRHL
TETFRKMLMSVLTFNSPVAYPAADARADEACDLVTRATVDRLRKTEETVLADERF
RGLLGPVAVARLUKLVIPGPPDPTARSIGLVFDNIRPKUSTEARTLGGMLIETLRQY
SGYLPVAVLDLTETVWVHVHEDTCLMDKRGVFRALGEPISRSQPLSSPVARTSE
STMKQALFDDYMRASNNKRTVTAPOVVAQESRNPFTRELDSCQLSRATENLI
QLTGQNLHQRTAIAESSGLVAMPWTVPSTNTALAHMLATNGAAIQESMLDDLSNRIN
DLTQSRRTFDJAKRVIGLQSRDFETLAVHVGLHPDDRELVLHALGPTSRSAQIDP
PEIHLRPNKTLKLSFTSINLPEIELLHLWTITWETVFCGRNLFTRRKHALQYTTT
GRHNEPSQSIOFMFVDSINAPGEIMDLRLYSPKYTKLIFTTELAVLSEKVEYVETLPEIF
AAQIGLNPSEVRILKNDACEPREIYVGNHELVDGDRFVAMEDV"
44891..45256
/notes="ORF 38; NCBI gi: 331248"
/number=38
/translation="MSVLCTCCCKCPNLIYAMGFFRRFTKTVQVVDILSANDPLTP
DISMRADVHYQAARLTISICQHPVAVMGTRDICALKGRATNIAAGNALMQLAKRWTM
DRAENSTNLVGIINDINS"
45306..48677
/notes="ORF 39; NCBI gi: 331249"
/number=39
/translation="MEATARLDNALAKISGGTFFDSCLSHGKVFIDNALRLBQOF
TGSNVAMETYGSLGPNRDFGQKYVIVPLIKTISIGRYLNRQAQNLAAAFRLISSTF
KPLASLPQATFTHMLFDITEORVAOTGNLQTVGGGTTTTYDIVMSLERKIVTTEIDS
LELFTNDPLVIAESITRELESQTLALDLSVILEHAKFCARQPLLEGMCGLREARQN
DLTDGAIIRACQLENIVGCSINRSPKLVSLHLTALKLIDTISPKDAAVTNTVFQNS
SIDRRLVESQVGGSDVYIYSWNGAVGMQPIAKFGKTPPGEGIRAFNAPVPEVSPGV
ARSAYGGLKSDVDVMAIHLGNNETIPVIPLDSTKESVYGNSTVETAMTQECIKLEFT
LVGFCPNLYDSLEPGSFFGSSGTGEEITPAMVNRSPRSTYIMDLDRGDRQVEITLRQ
LHRTVLDKQDAFDTSFARKIARLLEAFNEAEVANLLEPMEKKAADPGSIFE
FDRQMLLEIPDAADDVVLVGVDTMDRVAAYAVPRICMWNSSLSMKENEMIGVLS
FDQGVLPYLSDPYQCLAIKTFDDPAITDVGVRVNVVTVWNTIRAGFDALVNFKTI
FDEPDSFTFSVNNMKSICVTOALATENLEPMNIAVAAILPITDITLASMGGMVDT
ETNFMGVPVYRNACFICSGATLNPQKLMELTGVTKPLIVRAVWKCTREIKHFTG
QGNFMGFLYHLHLSTLIDKTVILKHFDVYKYSYGLCVTRTFRTHGHSVJPKKSA
RFTTGNMAYGMPISECTGKHAQTQCMELAVGPAGLDRGLGVIPICTDIHGMGVDFN
PRMAYALIFDANGFTPETTKTASPIYINILASGRYVEGIGAPDPVTGLSOCPTL
VAGLGDFDRLSLVRHLINPCDRAHYMLAARDGNVFEETIPVIFSNVYMRDLLEL
KRTVDAITADRNLYRETMLGIFADNYVIGKNGEHRANTKTVPPTTAAALSPKVLGES
LYNGAIGDLEYGEFSKLAKYTGVTTSIRIDANN"
48745..48750
/notes="ORF" s 38 and 39"
complement (48788..48793)
/notes="ORF" s 40 and 41"
complement (48880..49290)

```

polyA_signal

polyA_signal

polyA_signal

CDS

CDS

CDS

polyA_signal

polyA_signal

CDS

```

/note="ORF 40; NCBI gi: 331250"
/codon_start=1
/number=40
/translation="MTFVPPSSLSRIAPLSTVFLFTDNVRSNMFVCKIVYRCLSPRR
IISPSICTAFCSWANHGRLASLTAMRRSSSTFLFMASTRMSSSTVCFIFVFCVSR
TAMVWFSFSESTSLMCIISSMRLCFSTSGST"
complement (49423..50325)
CDS
/note="ORF 41; NCBI gi: 331251"
/codon_start=1
/number=41
/translation="MYLRDFHELAPSOTGKGVEYFVVGFTAGLKLGRLLVUNKRYL
IQKRGCVTIPNKNMYNDYQAFVVDHREYLIQIFFRSLPEYVINOGLITHRFPQVSGF
TREICGFATIRNLVYSDRSEARLQGLQRLSLVDHLKTTFPDDYLDLRLPLPSTG
GLYTTSTAEMFNTGTFDPFLAADFALFEGGCTHKPPIQFTTVRFDKCEGTCEADH
IHEITVLITGEVRRVYISNAQKIPILLVVLAVITAENRGVEPVRVGARPHATGN
IHFDTVLSTVYVQMMYLKVRVCAP"
50324..50713
/note="ORF 42; NCBI gi: 331252"
/codon_start=1
/number=42
/translation="MASFGERMSRAYTMSMGMECYKASTSSQRRRGVRDDTGTOP
HGLVRPITPPMPTTHRRPVPVVGDRGPPRRPEKEPSTTTTKKKGPPTAATRTTSKKI
VKNKASMDLAGTEKELEDECSPLFLEM"
50747..50752
/note="ORF 42"
complement (50977..50982)
/note="ORF 43"
complement (51032..53707)
/note="ORF 43; NCBI gi: 331253"
/codon_start=1
/number=43
/translation="MDALRNKIMGDLITGAGHMDRDFS LHETAAPLPIIQGPKNMKIC
AFINPTEKGLIKVYFELHGFLEKIGTEAEALRAMVGTATVMDSDVSEVYVTPDDSV
VSQPLTFELLVEARLLDLVAASVEKRVKICGGSTGVSTAVVLARLETRVSWPHNA
FLELAQYLYVYRNYGLGDFLALVATSLHLNDLWPLAVTRKIIPTPLFDVTVKQEEVE
MEVNGQWLSQYVYTFKAFKGVLETDSEGLSEDAADDPFLAESITLKLEKNQWT
KMDRODLKLTALLSINDENPRIKAKNNIKLUVDFMGIDRMAQPAFVKLAKLMGS
AHMVALQSTLRSSILSGNDVAGTINDSPDELLKRLRQTVYRMLADLSLKVR
TPMGVESVDKSVKLDLDRYMGKKDATADQDKILQFYNOQGPPTTEIEMLNKIGIESLE
IPEKREPGRVFNCLMSVYNSTETETGTLTNVSTGKGATSGTKNKSCLSDLGK
FRYVVRATPCTPPASILGLMLHLIRTLFGVWCVS IHMA GEFFAEENIDLLELNNKAA
TVWRKMGIDQFRGYFNFLGNDNDKKVKSDFMIFRVFGAKRALLSKRGNGNLRRS
NAFNGLEWLKVKVKAPELANNMRRARIINLERYDVQAMILLDQAEIRSLYLRLPWEWT
PFKTELOASTIILSKKADFNNAPTLLHVNLSVTISPTNTHTLPLAPIFNRKWDQHVY
GIILENNVTVYSQATGTATAGGQFGADYGDLSLMSLSLSQVSEDEPVTSELINM
FSLDKRVYDVVEIRKALSRYEETEPVYGEDVLVLSRELEQSDLEDCGTENNEPILVLQ
PFKDCSRGRAAATTLSELMWAGSTKIRIEGDDDED"
53657..54664
/note="ORF 44; NCBI gi: 331254"
/codon_start=1
/number=44
/translation="MGTADEIHPDWLAECIHPAHKKINGLITONEKKRAVFKSLKSE
FSETDLSIELDPYVSFRKYGLAADAPISWRHRLITLSFARFLAREPEASQYI
TORASTAVDCSGFLKETCAVRMKLYLPRVKMIGROGLRRKFFQCKKTPSLTPYKIL
LOVVVDIMRYTVEDVDLIDLVTIYPEFTLEELAKVKGAAINLEHMLIYHVLNAHI
NWFANRHFQILNPGLTIPSSLSQELAGWADLRCSCGRLLSIRSLBKTNPKRMKI
YYSYDSHGLISCSCKASVLKAPLSIGKLTALVINTPFFKQYHICACGCEAIVTNDLIPVS
E"
54642..55364
/note="ORF 45; NCBI gi: 331255"
/codon_start=1
/number=45
/translation="MTFQFQNEWNPILSDISDIIVDRTAAQWPTATYVGLLEADARAI
RPVQQYVNGPCCSLTWAGVPGVITQFVPPVTPNRYRVAAPGCAFGCEWCEPPT

```

```

VSAPPPSQFSDEFTSPELAPAVPKIDHEAPVATVSSPTSPRITTESRSVPTKEK
WGRKRVHKKTAEATWIPQIILKRGMSLSLLPAAARGQRSASVRRRVFTTNSKDVLYYD
KYSAPVMIQNTVTVVEKTSDK"
55360..55365
/note="ORF" s 44 and 45"
complement (55397..55402)
/note="ORF 46"
complement (55427..59494)
/note="ORF 46; contains 23 potential N-linked
glycosylation sites; potential glycoprotein; putative;
NCBI gi: 331256"
/codon_start=1
/number=46
/product="glycoprotein"
/translation="MKRTMLAIIILPLVYVAGETVADDEVFFLSAPQLDGLTIDMSG
TDAVITGSTLAWRLMKFTVSATHLREYGRVAKSVNSTAAKTCIHLCPVGTILKLSV
QATRSNANNASMRFAFPVITNHELNETCIPMEYDCEVISEKTSINTLSITTIYGSN
DIKKAIGENSTCNKWTQGCALATALTAYENKSVDAKKLMLRGTSYITCLLGLTIGIG
ELEPNSPCQTQGLFCYGLAEQVCTVAGRNPFGVPRADRPVSTGDCRVGSRADIQ
FSLGTLKGLFAERGCDDRVFVCDLIDPYDPATAAEFELAVECSRFSRVSGPGEDEM
GLELAPNFGLEGLINTQDLRLITITGIDFIRNNKNALVINHPIPTPELLDTFHFMA
ELDTERKLRHTVRLGWITCASGFVDFYENFKLLSGVDMVYIHPQIPMTSTPTLTC
PMPIRSDSLECGEPHRQCPGDFGVRHLVATRFLLGHISPEKJTLGLDVSQQLAGRRV
RQDYNWFDQNLFSVQGVGWSRGYITDVIQPNWTOALQINGCHGVCFPELDLILPS
ADATPYVGLLWSQSDLEDAVDFSMALGTRFSIEGALGDVYMDSSENIKTSTVHHA
VTVRMRAJANMEELGEGLAADFESMIWANVTRSTGATTGARKRRALTQCGDGVN
RTIPFSASVQSGTSSIVRGTCPRWCPCGLIASVRGLLESRTSYQGYRLTFFQNLVYT
QIQLNLGTCAKPEDLPVITQATKVSFEADINTFEGVNGSEYVYVLEGNDLQVTS
POVKSCAFINTNETQNTFTTIDERFFTPGRVADGFVIPAGINFGSONIMTVDY
IDENVTCHNYSNPVSQSIATVCAANVTECTPRATILLCNQATAAILLDFORSNELLKNS
LVDLIDLEHKVKMFAPSGSVPTSDKFGLSVAATITMSAALVASTAALAVATLAASK
IDGLAQADKDTADVITELGDSIALISAKLDNRINRAVGRVDDIQNQINQLIAMDNTF
KRLATGKELGTTTNERLGEVMAVYQQQIMSLTNQVTOGAIOIGYKVMIRTCVKS
LLAGNMGCTDASFRDHPGLTFRTKTRALLRQDKLFTVNEVPQTLTAKSVQVFIIP
SPTITEKVCWPDYKLMHVDGKVVAPLECTGKCYSEPTEATDYQECITNPSCTRCYVG
DCYRGICVNRITTEDISIKFENVTHALSVDLTSLPENSIPQIVSMEMEIQDILKLELIQ
LQKINTSVHMENTIGDIDAMKATIEYRAEKAKLVRTGFGEMIKYFIYAILGLVATIGA
LTAIIFMAVKCYQARALLSMATQVPTPRPMGMNY"
59028..60584
/note="ORF 47; NCBI gi: 331257"
/codon_start=1
/number=47
/product="serine protease"
/translation="WMNERVNEVFSEMTSSQSYSGIHVSFNSCIFVITGRANLHDA
LLLAEDRVACTSEFNTVPTGTHRWHLVAAVELTDFATRPYSLSVCVADTVKVFHNLFI
SVEPVTASVPQESISVPSNCGALKRNTSSSATVSPATATRGIRMTASIVFEITVLVD
GVATGSPNALVDFDCRFEMKHCVCQQLISGSTGANALCSYKGVGDFIVNFGNCPV
DWLPSTAPSYAVVSESGCYDDGAACDITTFGNFRDHRGRVAREGWDGLGTLIRPAG
NGCPIDDCGADGFTQAIQTVTVTYDTRSERCAAVLTVPPNETVWYDKCGCFIPS
SSSARAPPILGNMALLIARHPTLIRHQRILVRAAKPVTVYHGRGCHWLLNRVDRW
THNFGFGEVSPSRLEIARRELSTSRAPVAMSTLDTCDLSEVWVRVRLGATAPAVFR
GGVTVVEISPSGTTIEILLGRPLDFSRDEFGEEVTPLFWGPGRGKWTVCSTCGVIL
TTGCTCHGIR"
complement (60645..60650)
/note="ORF" s 48, 49, 50 and 51"
60753..60758
/note="ORF 47"
60771..60776
/note="ORF 47"
complement (60812..60817)
/note="ORF" s 48, 49, 50 and 51"
complement (60845..61681)
/note="ORF 48; NCBI gi: 331258"
/codon_start=1

```


CD5
 DKKWDCDLVFGTHAAQKAFIKAGNSVLGSLGGNYTYGDHVSNDGNTVPTQNLML
 MCSHYVYTRKTDVYAPWHEYVDQKRDQGYLMSLPAIIPRCKREGAFDIETIVHENAM
 DODLQNCCKFESEKSESVQLVDRFAGVTLGPPSPVGIQOKLHEIVSISLWVC
 NYHKTGPKKEYYYNTKMKENPHEMIPVEHLHDJASRIKFEACKNEFYHLLAIINR
 LRKSVNVLVYVNAQFDIQVQORLYRYAFKRCCKGHDDIPHEKGAIMEKWEAF
 LSVQFLKRAQIQLMGQDILKANYKLLLEGISVLAQAKSITMAKCTTKERIDSYRQWK
 DTQNFKSHGFCGDIIDMNVCKKEFEAKGCSLNTVAQLIIKFKPKHAKTIHKMD
 DITDKLDGTYRAGGTKIACELINLIDSLVIRIAKULKPMEEYIYRQLACTNIDTA
 AHTRGVNTCCGIGSTKWEVSRNKARLDAGIVMATDYINNSLFTPTPIPRGGEVMA
 PLTGLFPARTOCELCIDFTSMYPSMCDLMI SPTI VDSKTNRVGDYMGYDMSKI
 DOGEFEKTLVAVDRTPENPKLVHRHSTSLSLKRYLRIRTEHKRAIKQSGSVAEY
 HNRQNDKMICTNHYGVSEHTCSLMIITQOQHKIKLVNEFIKTLNRTGHSIFPNYGD
 TDSYMLYHPDSESETQLEDMVTLEDEMRAELREYMLKKLSAELVNRVKEKTKRTDTFV
 QSFLSDEVTVLFDHWEKLRILFSQGEVIEFFKDGCGTMMVDPDLTGIMDCSTPFSSEL
 ICKLEYENASSIGCHVAKKWSIGSTYLFYFKKISLYIHVRVMRCADYDGSPLHYFPV
 SLSTRAKORGDBH*
 76809..78608
 /note="ORF 58; NCBI gi: 331268"
 /codon_start=1
 /number=58

/translation="MFYEGCVPTLMPVLPITSIIFYLSLVHNLNGEIIINTKIKRGM
 TGFSSRFATESITNDFDLIFNGGALVKTGHEALKPVTWAKLSAPADILHSEPP
 VYENGICLANNWSITLIPHRVATVRLPCPDIGKGLVYVDLHESITGKSSVLIVLAD
 GCALNHEIFSSANVLHRELSFETIMAFRAFFVAGGLDPNSIVFYERIPRUKPEFLSKI
 TNAHYLTGKKLIKIPFTVLHKKRTTIQAIKVEDKKQIDYPEEMIKQDPANAMFRYPID
 LRLERIMDYFGSGLKCTLATPTPKYTVSGERSRHSIVIKNHDVKHYLINDSKFLAH
 VMDRAMPPTCYIHDDIDVKAMIQSVISRVRNIQDQVRLQELISAQNKLFHWFNQ
 LPPEMNNIDISFKPNLVPHALQOGKVPGLQYGNPEQIYDTMITEMTALLPRIGHWY
 EFMQSGELSSSKTLCILLPKIARDLVNPVLVTSLNDDTCGNMLIAFAHKIYIATAML
 HLLNGENLEHDLCPQGTGVVGPPEPKHTGQALREWFIDLQKTLADAPSTHACGA
 DFWLTHGSDPNFEKMYRVNGGVYRKRKIKCL*
 78587..78592
 /note="ORF's 57 and 58"
 complement (78633..78638)
 /note="ORF 59"
 complement (78703..79740)
 /note="ORF 59; potential membrane protein; putative; NCBI
 gi: 331269"
 /codon_start=1
 /number=59

polyA_signal

polyA_signal

CD5

/products="membrane protein"
 /translation="MVKGGLPVILKALKQLEGLSKAASVGTAEVIQKAYHQLTKLQAA
 RLVEAYTVLIVLYIMMLILITSSVIDLVGSETPHRLPPFVDOAQILKHNTAQDTLIV
 FVALGVNFLVLIVLYLILIKLKHIEAPLGSMSYFAHPVLKYIFIGINSFIFVFIIT
 VFSILRTICADANTLIVQDLLELLSNTSPADTNEAANAHSVGLLNTCTAPHTDIAKGC
 YSGVHLIAMSIESRTRRLWTGTSGLTEAGIPRMLTAVGSCWTKVVPVILLWFLIL
 YNF LHLHLMVIRALARRRRLGTITIEDEHLPILTSYEDGEDEEGDANLLYALPPGSTIPG
 MKKNYTPARA*
 79779..80960
 /note="ORF 60; NCBI gi: 331270"
 /codon_start=1
 /number=60

/translation="MMLISKARCVLLELLEMP IETINGDWRKFAKSPHVTTRCVTEGVALL
 HNSVVRVFGKVSVRDANDFLSFLYHDSIRI SAGTCEPRESMLAVILP IDKVEEY
 VAKTRRVEPTMGGFREACRAIKLESAAGVTIFDRIMKDIFNRSPTLAMGDGIPPLAKGL
 CVPILKSRGITPRGGLVGEIPKAGDHCPCKCAQYATKRSRFEVIDLVGDFSDNT
 CVLIEVNTKNSVLPILAVLKLYNTQTYINMFLFGIMYPHLRQYTKSLIAYVSPAGV
 QLFNVRSPPIITRMISAFPTFGEYCPQMRMTAAAMTYVIKAPFAAHTVGTDNVSI
 EVFSQSNKLPNPKWATDDAAREMERTEKARANKRANVADPDESPCKELRGEKSS
 T*

polyA_signal

CD5

complement (80525..80530)
 /note="ORF 61"
 complement (80932..81891)
 /note="ORF 61; NCBI gi: 331271"

polyA_signal
 /codon_start=1
 /number=61
 /translation="MISVARAMTDDSDSLWSTVDNRNSDMKEIMPEKVYRGCETHRSIP
 LLDADAFNERMLCQDDLTILKEGLDFRSTPFLKSFHITTEFGLVWVTDLTETVARI
 WRANINRHVGDTADPTVAEITRININHSGRTMMIPSIYSGPKNGQYPIILNKHLELFTL
 FATGKISTINHEITLPEAALLHEQTHLDRLDRLDPLFLFLEYEDGPASDPFFHVDRLGK
 LRARGEVMAACVETPTVPRSDVLMHMSMAEACVYLLQRETPRNMIALMDFIHVK
 TGTSTTMELMGMNRORTSTIIDILAFELFELTAGDFTWTFWALV*
 81073..81078
 /note="ORF 60"

polyA_signal

CD5

81515..82819
 /note="ORF 62; NCBI gi: 331272"
 /codon_start=1
 /number=62
 /translation="MCRVISATVGSVSPRCTRLIAHRIARIASVRSASVTITSPNSVS
 QNDRFNGVDRMKKSRENFRVKSWINMRSILKAASSSGIDLCVQPKYTFSGQMF1
 SEFLSTVDTTRESVMAARATEIITGAESTNKLGYCLOHELVDGFEFNRINILVASQA
 IKORVARAEKTLDELHDPVKRAMKLMGIGSDAVERMETECERVCRDPEGKLRFCOM
 ITSVFNLMRDGRGMRLVFOVELMRGFFIGIAANQFRDDLFKYKHIIIDLLGLATPDI
 VAPNPPCSMDTGREINRIDEYCKNYTLALAPQCCKTTIMVILLAAAMISYLDIEIV
 VQAQHKTCETLYDRVELVJLHEIQHSPWTPENRIVTINGTETREFIYDPAKGTTR
 VHF LSSSPNSVSIVHLLHSLINRDRVCRGRAVAGVCCRCA*
 join (81515..82717,97459..98016,98746..99543)
 /note="putative product; putative; NCBI gi: 331210"
 /codon_start=1
 /number=62

CD5

/product="terminase"
 /translation="MCRVISATVGSVSPRCTRLIAHRIARIASVRSASVTITSPNSVS
 QNDRFNGVDRMKKSRENFRVKSWINMRSILKAASSSGIDLCVQPKYTFSGQMF1
 SEFLSTVDTTRESVMAARATEIITGAESTNKLGYCLOHELVDGFEFNRINILVASQA
 IKORVARAEKTLDELHDPVKRAMKLMGIGSDAVERMETECERVCRDPEGKLRFCOM
 ITSVFNLMRDGRGMRLVFOVELMRGFFIGIAANQFRDDLFKYKHIIIDLLGLATPDI
 VAPNPPCSMDTGREINRIDEYCKNYTLALAPQCCKTTIMVILLAAAMISYLDIEIV
 VQAQHKTCETLYDRVELVJLHEIQHSPWTPENRIVTINGTETREFIYDPAKGTTR
 VHF LSSSPNAARQQIPDFVLIDEAFAFNPAISLLSLP LLAVKRRKQIISSHIAKSV
 TRVETIMSGDKRAFHVINQRFKCDGHAPLPMWCPSCAFCTHIDLINESIQELINN
 ITPGCELELITGCGAGVLKEDVATPFSEELVRFQFHTNAPVDSLGVKRVFIAVDPTA
 NGTQSGFGVCTFAELNNGHFLVVALEEAIGEMRSIMHFYTVVLAHAYTILGGINR
 KDPIIFLEPONTNEDNETLMSHLSGAERKFCIKLIVYRQFNKSGTTEILGVLGSGN
 KASWIKVHNMEADMLGRVMTYASFGELVAGLYLHNKHVLSKYRYADVTXTIIDARD
 AIANPRDIEDFKKLPADSAEYIFRYLGGTEVSVAHGDDGIFKAGTELVITRIGCELMN
 VSLKSVGRTKVITGKRYGGSYSRDDILFSAFILLGTSMMGRPDATDGLFVT*
 82758..84746
 /note="ORF 63; NCBI gi: 331273"
 /codon_start=1
 /number=63

/translation="MYVVSVSEPLLAVCWADARDDAVIKHAIKRWICATVDIQPVTL
 GAVFFTRIPVSGRLVENSDFHIVKLPIDGQAQIDTDRMIALAGTCGGEANVPSL
 LENWFLGFEPEKTATMSALYTMESALYNHLLBKGVDAFHIIIFERRFCOOLHILGPV
 PIDVFEKWIIPDALLNERETGLSYINLDKRNFTKESRPRLGKVLKRWLADNMIKTN
 PHOLIKYCLNDPDDLVSUPLPPNALPHVPILIMDDSMHSGEVFOYVITDYKSNETV
 PPGAYKEVLEGGQIVTVINFLNDRKESGLLDPMCAIDNICEIFVAFIHRHVEKFEFNFN
 RVDRNRIGEAVFVRAALPGKTSARFTWFPAYELCFQITREADFTFQELIMLDE
 SFVYTVTNDCGVSTRVCAIDAOPFCRNKSCRLPNSTKMEAGEPRGAFEYIKSYNALVK
 SNRGYSKNNIGISRSFVAFSLGQYIFGALARFASNETVYNSHASENITYEKDH
 IDNAKYLKLTWGDILKISPTTSGSVRLTPWMTDRDFCLVHNR1HHKAGVSVIVTNRRI
 YPKCFEHPDPPAHVPIEGGFLDWVEVAGKPRARVFCGQ*
 84664..86208
 /note="ORF 64; NCBI gi: 331274"
 /codon_start=1
 /number=64

CD5

polyA_signal

CD5

/translation="MSSPRGASSMSRSPVNLFEPSDSDVSEWVFKYRLILGALRSAT
 LEGEDLEGGIPQELLNVLVLYERKMSTKSELSGFVYIALMTLGI DFDORSIKRLVDE

GVLMGFRNRMTRVFGVATGDDALLIEGLDTHPAPNVLARFPVPIAQLIDSNFCV
LCMYSDHSIATADTRFKTILCOHFTSEAVTAGDPCDCAEYATVAKKTIQIDISLDV
FHEARVACONRELAUKROTTLTRKNGSINNFKAUNRARETCAMDREMSLSEVFOK
KLRENNMINRITLTKNIMLFAELFVNFVHAHRRKTIINERLSDGLGCKMGRAAHML
FUTTKTATVAKNILLTTITTEFAAAYPVYRNKYNHIAHRIFFKHMESIGFTNE
GITAFQOLNOVDLREVVGDYRSDIIPITTAENENRLETFEWFIANICTFIFHNK
TIKLHGADPADIANIDELPGLYFNCSVGEKIDNGETRSCWSSERTILDLITDYLQLL
STGVIG*

complement (86254..86259)
/note="ORF" s 65 and 66"
complement (86367..90671)
/note="ORF" 65; NCBI gi: 331275"
/codon_start=1
/number=65

polyA_signal

CDS

/translation="MEINAAIESLTSITDLPDTAVCDHENTYLRATMDGALGMSHI
KPVISEVTVALEDDTVGSRYAIVTAPRYKIVQPYTDADKGYLRGLRTPCAEWY
MHIYNKIISHAPANLVRQAVOSNMILGLHRDMLGYSVDDIYTRLDVPAPQIVIG
GALYRILHSRNALPKCAIGVGFQNMVDTAQPCATPEVNDILNVAEYELPOR
VVAOLLIEWLETSPVTPRGTRNRIITLELOKALATHLPAGLFHDIVTGLIGLPPRA
VSGLEEKMGIDLFNKTQNSYHGRDVSFRGALENAARUCSVISATGKTAPVTFGL
VTVTRNPPFKLSKLVTYAPHPITRSPFLCARVRDYYDOETEVPHGRASGPRHESR
MVLREAAASVPVLTNTLHESFSGFYIPFLREYKSIYAFWSTGLPSRSHLYIF
HVLRSNIETLTPDPVFPFKERVEMQCCLAFVRETCHEELYALKIFRLUWBERGYL
PMYARGUSAPVQILGFIIVNICALLAFDDTDGLSDSDMDYDLSALGRVWVWMLIT
TLFPLMYTATPTEFQCTYRGENYVYTTGALYPGDNTERLSVSDYDGLADE
KTKGGETLINPVHIGIVNOQIFPILSAQLPYARELPNKSVRHPIIMIEAVRRLK
SERFGSCVPPVPGVITP SASVPTGKYIHVMKHEAPYEVTVASNGTGDGSLSDSD
DESAEPTSPASWIGDRDEKRLREMEATVRNDLILGGRETHTGCPVTRQRPVR
APKVDVSLMAEAKRSGLMDLDVIVAKTTHMVTGCMFPVDTLETLPDPPAD
LLQTSVDSERSIAPGADPVSHDLKDLPLPAAAYEEDPLDLPAAAYEELDLDLP
PCAAYEEDPLDLPAAAYEEDPLDLPAAAYEEDPLDLPDLPDLPDLPDLPAAAYE
PRAGSESDGTGTTGADGSLPVSPPEPMSAPGEPATGTTGVRGFKSERLILPRKIMS
SIPRSDALKERLSTRTOMYDGSDESSEYDSDSDSDGRTGRTATSTLIDEVER
MTHSFNWRISNTPVATISDIPALDPPSDSISGSGEGAAEEVOLGADPGGHTRAP
PVADIRGALDRYLYKESLSLSVIDELFTXTVTPAAPPSPGVPTPVADNGGTRET
HPREKCPYTAEKYRILYTYGEDELAALVESTGAGTGVRRVRSARVLANVPVSSAVI
INESLEDQFTRILENTYSGIKRVIYIPGDGNCILYNTLAPTAGADGESALDKFELLD
DIRKYVRNODAEERDLILTEIDNLAGPNVYSGDILISFFQLLRGVGVTVVSMKIGR
LVKLVAFTOEGIPPEYIILFTNSHYEPVLTEDTHIPDSYKDFIDKRSILFSLAI"

86406..86411
/note="ORF" s 63 and 64"
87758..88025
complement (90662..91897)
/note="ORF" 66; NCBI gi: 331276"
/codon_start=1
/number=66

polyA_signal

repeat_unit

CDS

/translation="MDSITLNAFVSSAYTGEFVYDLDLNLFNKMLGGDPAPKPTQP
EERTTRGPVPPRRVRPRAEDQLEELVKPDIGCPRGEQVEEPVEEPVEP
RRIPDPRNSTHRELSADSPPSHGCAVDFDRBPVTQKGRPVRDMCPVRVSYDI
STDVDRIMANLQSWKESRAVKVSEKDVSCHEACERKYVGAALMTVSALLRLAWP
DKIRIDHPQIKLLYNASRDVFNVLTSQKRSISGKVTNYYDGTQDPLTNLGLDNR
DLKSLTYKMAOLMKYETKWERFSAAAGGYDMKDLIEARLRVERPFEOTSOLVDY
LLDSIDDLITIGERAKLSDFWTLISRLKLHQLLWELTEGYLKEENKVIKERLD
SLQVQVRKLNQRNLDLKKW"

91604..91634
91672..91693
91887..91657
/note="ORF" 67; NCBI gi: 331277"
/codon_start=1
/number=67

repeat_unit

CDS

/translation="MESIVYDSDFSDDESVALTLDNLNLYTRKRIISDRRLRHFGGLT
DYTGCTILSVPEPNVVKWPLIFPVKEGVLSVEIITAGTVNKNOLVDRGCHWETTCV
TEP SLPVSNITIGIVVRSRIVASGESEIFLLLTDPGEGIGFKRMPCTINTAGTVCA
KISREVITDAPVWLYDDIYSLPLTASLPDRVFLDCTGWTQAELERFAGHLGSPVDM

RYTPRKILISPPHYLTNNLTTLETWSDSLDLPPEPGFQDVFIAAPSPKIRYTDGMS
AIKRTDTCVWTHGEEDVTFSEAFNRVETLEFCVADAPVNPQMSLVSDMTKRVVTE
AAIGTLTAVELAPDGVPEPLYSAEYVYVTHDRELRAYACVSPALSTIVIGALGRGV
LHLQLHVELAMTGMFDRFMKFRREGNRELATVDEAFIAAIPACAEHMRVILSRGH
GSRFLWESGSHFRFLMADPMSPNPRASNVIIRPQGVSLPPRRNPRIIYEWTA
MVAVTARGALVSTRELTAERYRALSITVNMIGDMLYIPVEPLDGGGERLIDLPLF
TGAFQHAEPVVDSDVNSAGSRRAVIALHRLDYLEDARAVICSNVFGTACGAAL
QVAPADLPQPGFTVDSLVKPTDNPDKVQLQTLHRRNGFCGAYDEAVTTGFLGLSG
ERMSIKOPSAICLMSFYEKLHPIFLHGAKIKCYVAHTEHPCTVDFVGCIVATSHMQV
ADNVIIMARNPSSATCKTFTIDSFNESHVRAIFKPEHNIERYAQLEGPAYMENTGH
VRENGTGREVVKLTGCTIDTRTPGVQLVTTIDGREFVITPLGESPLMVIDGRITMID
LTAPGILVSTTNNWVAENSRKMTREYSGMSYECYRIEDRSDDLSCPPQGRWR
TGPRLDLRDPFEEHWWQAVNDGHTTLRNTGAVELFEPGLINTWTSVVPVDRGV
PVGATGAMVSDAPMREQLMALISDTLSFKYPRANWSENILSLHGAILKMQTVII
TRKDDPTIAHGGVKITVKEPGLINDTELPRGIYERRQIPWAHLHVLVTFRLLEGT
LLIATTLRVVHVIVDAHAGEQILKMALEQAPLFNEPKRVVITGLPSALPATIYIG
AMPNINIVATGQGLKNSVRDLAGVRYVSGDDFDVGLTIDREFFYAITGTGRAVE
SFMRLDPAPYIVELAPEGSLPVYRDPFNPICEVTRGFNFEVNPPIAILTDSNTVM
MSRGDIITPGERRVVDLPYSAMTPERLOEPPPTTFLGSGDVAIVTLRRRPPVGTRT
KSLRLTRNGVTVMLQRQDIYKVRVLVDRKAVKTSGLLVTSHSIGHTDDVYVPVAGGEL
DYAKTWSFHGLGGEPCLDVVTVPADCLAVAGAVRLFVLSLTHPVTFCGWNIVIPRAL
RPFRTYRFGTPIVMDRVLYDISIDRLITVPSGDRGRGVIHETRVSDEKSRKLYFLK
QRLISSGHRKRYWTERLTGTLATHNRWLGAAWLVGCGTTRRLLSRENFIROGSRST
PQQQ*

complement (96355..96360)
/note="ORF" 68"
complement (96520..97740)
/note="ORF" 68; NCBI gi: 331278"
/codon_start=1
/number=68

polyA_signal

CDS

/translation="MRGAECGAWAHTKCVVPTKPLVDHVHVESFPFIATHDRLLHA
KDWPLGYVRGDMVNLFSVLIDGEXREKTOERRGVOKGLVDENEVDLSGCLIKTETIK
KNRLITGSHLSDSDSIIKICIQCGCTSAISAGINQCPIHFTVHVCHAVRNGTCQL
KRKIAVLIGESSPQFCFFDNGRADPEVEVRGITGKIITAGLDQKFEKIAIVFGHGDN
PTNITFKIKYIMLIRKFSFIDALAPLSFHNVIDIIOQNSGVTRQDRTKYRWETI
EEGALSHNVDEQVCRVKILSSELYHADLVLLMTYKIKNSRHASDDLRFYLVPEL
CIHKLITDKYRTDILLINFLEKKSSESQVQRMFTQFNKVLHRIYDAKLVDICVYVMS
EITALEALNVHIC"
96753..96758
/note="ORF" 67"
97459..98016
/number=2
complement (98034..98039)
/note="ORF" 70"
complement (98038..98721)
/note="ORF" 70; NCBI gi: 331279"
/codon_start=1
/number=70

polyA_signal

exon

polyA_signal

CDS

/translation="MDGYFKNITCEHIIQNDIRRAILDSAVEAEYSYRLKSTMTIF
YSTYQHYLAIDPHCNVDVNOTLGSNPKLLLLTLKLP CODEMDYIISLTLGPHITTF
IWRSTPCTKYAAVCATCMCLAHMAPLKRIMEENPELFEEFYVYVNRIMKNKFEQ
TEOCCRGGRIMYNI LLQCAQHGDPEISIFVGLRGVIAFVRLSTGLVHSRQLPVDIPNIDM
SLIFKHFVY"
99853..99858
/note="ORF" 62"
100000..104052
/note="ORF" 72; NCBI gi: 331280"
/codon_start=1
/number=72

polyA_signal

CDS

/translation="MDGGMVYSKGNWFFCMASENKRPKAAILSKAMGTOASFMTD
QLATKCNPEGLISIHNSDEALAVYSAHTDGLNTCKICGHNVDRIDAKHARQVHLGR
LCKIENLAEGLAQTLRIDRVEFAVYDCCYFWIGARAAPGLSRFPKPDATRCHRCPT
VEVPDIENGESLEAVRLTGSPRLTDRAHLLAFHPGDFMASITTLISPKKSNMCAI
YSARGELQRFQJMWRCYGYKGERATAPTKAVNEGFCITAVHEFLKMYDARTKALSILP

SDTKLYTVFMEGEPTTAVIPPGTHSVVPATTPAPAPCKLMSDPLELIRKGGK
 RVIMDSVAVSGSLVAGSPSERVMKRSERPLDSSAVKRVSKVTSVTDQECTAM
 ESSIPGVSIEARPTITYGSAQTTAGVLTEGGVAPYPLIMDVYGERDLPIDLPNOIIT
 GRVIESIGDEGDIARKEVIAKVEALVAYSYPIMQLSLGIVFFDEGHIYSVLEKP
 TGDILMSQITLRVLSITGTFPGKSGPAGTKMAIVHHIFRDVVKLVLAQNRKLVTV
 FKQDVYVWEHLGTLTAPNSVEPRVDITCAQNLIISSLSOVTFQHTTKLP IKPEVRAT
 LSPDVLKIDMPRPSPSEVTEMACRIGVWGDI GHPTIOFPLDEKIHMGSEERLDPV
 LLIKPKAAPSPFYSMEVIRVLETPGRSDESDWVPSALISHATLAASORGLVAA
 HASVGAQAALSIRGAGFATPGSGEDATSAIMKMIQELSNTSIPTSRSLCELGSSM
 RGP AVATPRPAEAPPSSEGELEFDVDAFLKWTGPADVEGTAPVVEIP IVGGDI IRQ
 SNA LAGECHELITETSPGNVATGETTPATPTIINPPSPHVPVMPPPANCKRTKS
 VKRIEASKEGGERDATGSSMOTVRGTGGLVTPKQSLIRGDELGLMKWVFEPAL
 DTRTSYLVGVRHTIIVORLQHGSKMKRCRDPPTFAITDADVP ILTTPDATRRRI
 CYVDNLPSPTWFTWGPNGSEPPVTEHRTSYELRYGDIQAHAHQNLILGPSMEIYAY
 SPRVMIAQHTEDGANOIGWGFPPYIRVYVNNMKIRPSKHIPYIYVIEPHSVGTYKM
 NYAYLITPMGHPVLTTCVAGYSKYLKALVRTLHVAHSAGIKIGYMHPRKNVWKTDAL
 CSDLOTYIFDVVTOILENVFMPGSGGRNAPVIPPENWTKMDNGCPPNLNLTASDI
 FCLGRLLTVFCAQDDLTSSQVALIQIMCAKPEERPCVTDLLVNFVNSPEEPFYPVDI
 GEMPVKITNTKKAERSNTAYVRFVPQAGE"

repeat_unit
 polyA_signal
 polyA_signal
 polyA_signal
 polyA_signal
 polyA_signal
 polyA_signal
 polyA_signal
 polyA_signal

CDS
 /product="protein kinase"
 /translation="MADRTPRSADGLIHDAKPSKVTKNDPPVIDFKRIIAAKIAEK
 RGGPPVAVSLIAASTVSTPAPVTYGSVGLPTROLSDSDDDDEEDNATAKTHSVIPV
 FKPPDQLTAPHTVSKRPTQPTTPRPVHOQAQTGRPPVSYDTTGRATTQDEITDAF
 GELNLEIRLALATPVPGMPPESTRGERNQVNPVGVKELPIVQOSITDRAIQENRL
 EAQFHGATMVKTVTPVTSVRGVFTRHVITDTRKQPFLEACEFIPSMIQEAPIPFR
 AVASKPFACTKLITPKACAPAVDILRAPVLCGVFGYVYKGNLACKVQHGVRMPGV
 SNAISDVVEESLLGSRLQHPNIIITVFKGELYHGAVNTEAVCVSWEIGMLDLSFTQO
 SFWQKQACDPLVKRYLARPEKHITLLGLEHLHGLMRHDIKSONIFITFNRGLV
 AKLGDGCKSGACFCADGCTRAYEPPELTIAINVOCASDMWAGVYTWVEHTGRTPM
 GGTDSIMQKFRYTGEDNRANQAVARHAMAADSAKVPDPLEGLMERKVSLSIP
 SFTDIMKSVIRINPNDRATSDLSKSPRYTDESITGECCTDRKSLEKNRPMIRLIE
 HNHLPPKVIWTPDHERDQAEKRGVMEVTDGFRPAPLYMPWILARADMGDDHT
 AKKLIITLTDULVSVVVAKEYGVKEIRVYRPPESPTDYDLYLEKPTIDFEAQI
 SLMEGTQAQKSPYVVPVHYTLGAHGKRYMIYVTPAKRSITELNFDGTDGDTLL
 GAVILKQLISLAVAFRNGINFITNNYNTHILYHDSRVDIGPKLMDVIYMLLHQTN
 NIANYKLSSTDRITLDPGDPVLKTCLAATYVYKLLMSAPRALERLVPKTTISACKRFD
 DFFMIPVVKRISIPDPTKIVPKLFSFKIEPPQDISGGTVYASGTLIS"

complement (108143..110164)
 /note="ORF 74; related sequence to ORF 73; NCBI gi: 331282"

/codon_start=1
 /number=74

/product="protein kinase"
 /translation="MSGKVEINEPLIVATSLRPNCTGGSHFGKTTIANRALGSSQL

MKCSLLKRCVREXIPVSDRGFLDAYTQTAPGIVFTGSSLVYDSTLMMARFNH
 RQHVLPLEENFGLERAGDVTVDITRDDEAVARGLRSDMYERDAVGHIVSKPQKRTT
 GEMILDEAMTARLSHGNLVRELGVAACEETRGGKI TVKMWELGICNLDEFAEPHMTA
 PKFAIVROYTICKLERDITSLGLELYLHMHVHGDGFRGAKVIVFP TADDETFKNLGEES
 TSLIGATPDSNAPYHRAPEFETGLIGREADMMAWALTMYRAHTCKDFKKGDPSEPY
 PRLAGLMDNDALDSRRTGDPNPEGFCITRLEKNATVGGCTGLMGKCLALYSPQFR
 PTARELLVYPRYATLHRTSGPAPPVPIGAAPAVHVSNAVEFRSGRWGVRSPAMPD
 RLEARTSGKNNAVAVGDFTPSFYREGGLDTRPKNRAITISRGTLASVYGTENRL
 TTEIHRSGADIVLHKVYGRRPDISRANLEYVMKSVNGPEMYITTDHYTVGVNTHL
 YVGYITDDIVASAGLSCP LRVDIILKGFEKSTLRVDLMLVTRGLAYTLVRIRHLK
 SASIYEAQGHAVGLVFVNPSSNRFECTVSCDKLELGDGEPQLITGSR"

polyA_signal
 CDS
 /note="ORF 75"
 complement (110249..110254)
 /note="ORF 75"
 complement (110271..111542)
 /note="ORF 75; NCBI gi: 331283"

/codon_start=1
 /number=75

/translation="WNTIKSSQCVHCILEAYPPGAGPVSAIDERMIRIFMPGGETLM
 IDRLIVYEPSPCTYHTGLDLVGAHIVPNNSFDRDVTGRDLSDFRSCDFMEDVHPW
 FRVLEITTDRLDGPGRGETVINFGRGVPVGVYLCSHGVATASCTIMGPVQGMPE
 TISITINPFGWFSWLVLKAPAGTSGSKSPFANDLTRYLIRTSFGSTAGIDRETFSP
 ILYGGRWVLCGVPHLPTLVAKESIECALDPSGIWVIMPALPVYRHLVETVLDVR
 VRLVCTLAANFEDIFVHRVDNRVTHRWVEIPKNAPRKLSTVTVGHRPEVLAAMCG
 GGRRLIGHTMPOMGFDITDLKVEPAPGVKLDGPPQLHPEFLAAMAKIEPLVAPIESEL
 RGNTHLAWITITVVALVGRFGILITALESLIS"

polyA_signal
 polyA_signal
 CDS
 /note="ORF's 76 and 77"
 complement (111526..111531)
 /note="ORF's 76 and 77"
 complement (111560..111565)
 /note="ORF's 76 and 77"

complement (111597..112322)
 /note="ORF 76; related sequence to ORF 77; NCBI gi: 331284"

/codon_start=1
 /number=76

/translation="MDRHRYLSITLTPVAGVPAAYELFMTERGSIATIPPVDRSD
 ESRVESIIRAHGRPIIAFAGRMGSGKHACDYCFERFGAIKLHIFESGLRILGIFYGR
 AIDKSDRSVLQAVGELGRRLPQFWLQMTFEEDLLRLILDRDPRVITGVYKTDVRY
 LATLGVRTFITRRPGVLVLAGTPEESDPSERDLDVYHECAELINRGTVGDFALLAETA
 RDISLSLVPVTDITGRCPRIFP"

complement (112497..113183)
 /note="ORF 77; related sequence to ORF 76; NCBI gi: 331285"

/codon_start=1
 /number=77

/translation="MARKGEPPSYIKALTPVINSVAVPEYPLFVYKMSGGFP
 ENSNRSTGRYDWMVYDLARTKKTTLIAGRMGSGKHACDYCSIKIGEKVHVFDAG
 IMARFTEIGRPTEKPRDPFLOAIGDLGRRLDANFWRDTEKIVAPMWSGVNVFIT
 GVAFPEALAGLGVRTLIRPRILITGSPMERHSETALDEYRDCPEILINDTAA
 FEDRVSTELP"

complement (113341..114543)
 /note="ORF 78; potential Zn-binding protein; putative; NCBI gi: 331286"

/codon_start=1
 /number=78

/product="Zn-binding protein"
 /translation="MELSILYELPKRRAVQRTQPPPKVHREDTGVLPGLVRESS
 VRPDSITNFRLLAAVKFNKPEWEGCGGCECKTRPALMNCDCDCTAKYV ICIGCK
 PRADVHPCFNRNTCVSGCTIGVAVWKLGGCGATCLDCSLGADWPKSRTRGTGCVG
 GAQTLGVHYRPMKI IKNTRPYIAPETPIPQVORHISARILEEIRRCQLSRSEIVSLFG
 DHGSMWEYPTAVPAHLAPMIGSDISKAFADDTILPGRSMAAGVPGWRGCGQA
 PVLPFGMDMCLGVRAEQETITQLVINGTIWEDAP IAGTYHMSWMLDPTVTVSPVD
 ISALDTEAVEFCNGLGSHPEKYRLSDITCRIRWTAGGFLDPESLWVPEKERLANTA
 SAHSNNSH"

CDS

/codon_start=1
 /number=79

/product="protein kinase"
 /translation="MSGKVEINEPLIVATSLRPNCTGGSHFGKTTIANRALGSSQL

CDS

complement (114588..115133)

```

/notes="ORF 79; NCBI gi: 331287"
/codon_start=1
/number=79
/translation="MDFEGCSLVTDGERVSNPTDRGFRSAAPRGMLKPSGRADGGV
WFCDDVTDYDITVAHGTLQDDNIVFMTRLFPGGLIPDDVKSDRHVCPDCGVKVR
FTLVDDFQDEHYHSGCGPRIRQFLQQRVSELYELTQFKGDRPYARRKWDLSRD
SAVLKAVAVYAPKDNVQTC"
115168..115242
115246..115251
/notes="ORF's 78 and 79"
/number=78
115291..115447
115567..115572
/notes="ORF's 78 and 79"
/number=79
115733..115943
complement(116137..116142)
/notes="ORF 1"
complement(116253..116258)
/notes="ORF 1"
116279..116418
complement(116412..119000)
/notes="ORF 1; Identical sequence as found in span
742-3330; NCBI gi: 331288"
/codon_start=1
/number=1
/translation="MDCIKETPAVASLAGGTTDLSTVWNFDLGDMDQASGWDIDV
HPVVEGASVDAMDVAESEDDEEPTGRDEPEVESEADDEVEDAIPAAVPST
IPLSTTIRGEGDGLGKNLTPADLAGNLGKYAATLFSEKTLPTLAAVVEQKALDGC
SSVAKSVTLISSELNALGIFSRPLTFQDNQAAVMIIMKVLGRLTALPLPTKTDGCV
LSAIVLFSKARKVILVSTHSDITKFIENDDGTIPTLGVDYVAPLAKYIDYDRLKLT
LVLIKLNNGTVKVRKSKTKLVMERIPVECFLEFRAPVTESDNIYATIDFVKR
LKGVSQSSVYVYVKNPKSIGDDFLNKTNTSLTGADPLTKVLSFRPNTATSSGFK
RWLLIKCKLARGGRIELGPKSTLRSGHEESLEEVEVTGCIKSMASLTRLASRAMSPK
NLIKANDRPVWTPSILNEIPDEVHRTGNTLIPWLLFAASLRDLIVANKWLESTEGCDPR
SIDVAKITVHYQVLEIRLLSDAGIAKGLSKLLTANMLITEMSGFVYMYKGSYGNPT
VLSTISDNVNPVLPISWIEIPLWSTSGIVKVKIVHTNSSILVRMLLSIDVTEVAK
VHAIAKAVRAEQKASKESGADAKKAPKREAEAGTETPKKKQKEEKSEKPKTK
KAESTGTAKVTKKEKTETKPTKSTNRKNKVSVAEAGTAVPDEANGKKSTQVK
RKESAGQHTTEKKRTAAKKTVDPRSGHRPSKREYRSQEFVSDDSDSDSEVISKPR
VTTEENKSTKATLASNMDDGDDGDMTCAGQFDIGGSGFLSTTAGNRKSNKGS
KTLILS"
119172..120395
/notes="ORF 2; Identical sequence as found in span
3502-4725; NCBI gi: 331289"
/codon_start=1
/number=2
/translation="MTCGRSHPRQTSPPRPVEKMPRFQEHGASHRYTPMCKRPAAP
ANTGSHRGALLGNGDPERKEETAVSKLILSKRGYIPEARTALIEWKEGARRGGF
EVMCGSILLPGHDSGRAIDDIFRDMRSVRGEVASPIDELFRVGNLLEVYVETTC
KEVPKAGKGVQVQVETVTVGIPKMIHVCLDRSLGVGKCSFAKLWKXTIEELLGTCVG
TTPHQCMTFRTIHSDAIQASFQGLNMIIEFLFCGRVTRGVRVSGERRVFTHRCLDGDS
FNAFLAATVHDPVTVSRVAFVARTVQTSYPAMYTFGDMPTPAPNNVGMVTTTSGMC
YDIRVRLVNDPNKMSVTPAAAAAPPVGPKEGHEELEADPHKPSSTIGSPAPNICALY
TSGDSVELHAPLILP"
120556..121464
/notes="ORF 3; Identical sequence as found in span
4886-5794; NCBI gi: 331290"
/codon_start=1
/number=3
/translation="MAFSTSDKRIAEILLGMSPATLSVRVFGVDLPHPDIPWDSDEG
DDTPTGWSERALGNETVARPHVYIGPVGICAVDDTSDWDSFASPVTGAEARAAGG
LPQAPLIVRRVSGNEMRVSESRRIIEIGRELARICEIQWEDVAGAPVGTGNS
PKSPVRCGLLPKSGSPVYNNPPPPSPDDTTCELPISLPSLVSSPVNPAPNYSFVSD
APLNTFSPNNMTPTHPGLARLDTPTPTLDLDGWDVMDWLYAPQLGDLQRIEAGSS
PKGRLDLDDPEVPSKSGSPSRRALLN"

```

```

repeat_unit
polyA_signal
/notes="ORF's 2 and 3"
121788..121793
121828..121833
/notes="ORF's 2 and 3"
121864..121869
/notes="ORF's 2 and 3"
121904..121909
/notes="ORF's 2 and 3"
121937..121942
/notes="ORF's 2 and 3"
122183..122956
123630..124169
/notes="ORF 4; Identical sequence as found in span
7960-8499; NCBI gi: 331291"
/codon_start=1
/number=4
/translation="MASIDVGETIGIEQVPEVKFRVCHGFCPPSFVTARGAGSVSL
RVVHYVPSPEREDPESPTRGVDVDGACSEPTPTPRPRERAIEEMGKIVIVSVCP
LRPALQHRVWGCAPARRHSDSVARRAFEPWRGRASRPTVLTTPASGSDSDVDTRTFC
CGCFGHPFVDCMCDQDML"
124227..124232
/notes="ORF 4"
124312..124317
/notes="ORF 4"
124455..125141
/notes="ORF 5; Identical sequence as found in span
8785-9471; NCBI gi: 331292"
/codon_start=1
/number=5
/product="thymidine kinase"
/translation="MALREGPTPARELPEGLVFCVEGNIGCGKSTLVKALMERVAGSG
VNVVEPVDQNVHNGKYIELSYDTPGYAVFQNLVDFSYNVQVQNLQNDIMERSP
MSATRVFCVNGSRGVIPATALPCMAARGEAVMTIATRPVFVYLELPPEELARRMR
RDRTEAGVGGLDYLRLLHERYEAWLSSAEDVERVDASRREEIVDRVTEILLCRHRL
RAPLTRKSQL"
125205..125621
/notes="potential product; ORF 6; Identical sequence as
found in span 9535-9951; potential membrane protein;
putative; NCBI gi: 331293"
/codon_start=1
/number=6
/product="membrane protein"
/translation="MNSLTIIIFLSGLTAYHAVIADGTGSSSESVTAGDSGVVWLMTIG
ALLTLMTPIIGLFGIVYRTRASIEEMRGLMQIHLRLITGDRSNRGDELGAGAS
LLTSSQPPSYAEALLMEPEVQPGQGVPLEAEIRV"
125713..125718
/notes="ORF's 5 and 6"
125918..125937
125968..125996
126090..126284
126533..126955
/notes="potential product; ORF 7; Identical sequence as
found in span 10863-11285; potential membrane protein;
putative; NCBI gi: 331294"
/codon_start=1
/number=7
/product="membrane protein"
/translation="MAAVILERAEEFVAPGEARVGPVILAEVYRALTSDHEMRAFYET
CAVSFFADLLIIVLHSHRPEGSTTRGTDHAHTQEGSTTRGTDHAHTQEGSRDQGS
MTEADDLTRPPLGCHGRQIPVIRRRVLDRLRIDYSL"
126729..126822
127169..127174
repeat_unit
polyA_signal

```



```

repeat_unit
repeat_unit
CDS
/note="ORF 7"
127328..127421
127437..127568
128132..128830
/note="potential product; ORF 8; Identical sequence as
found in span 12462-13160; potential membrane protein;
putative; NCBI gi: 331295"
/codon_start=1
/number=8
/product="membrane protein"
/translation="MKCPRARVDIFKREIKTRELERTKENVYHGLTGPISLPPHNPV
IRFQVMSRLRAGVIGFFLLVSGSAAISPSAPTDTATRESPTGCEHRDRALST
ETPTPEPSRGGSTEEVLHVVTGVPVRDPRDILIERLAETLSHOLLTPGTGPR
EUEDEFARALAAAEITAGSADVRVWKATLSLMLVTSVTAGVALWIVARHGIFR
VIPERSRDWSGPI"
128867..129343
/note="potential product; ORF 9; Identical sequence as
found in span 13197-13673; potential Zn-binding protein;
related sequence to ORF 10, ORF 11; putative; NCBI gi:
331296"
/codon_start=1
/number=9
/product="Zn-binding protein"
/translation="MATRPKRRPRDRALVPPKVDKRIKIFACGHRKKEICELGYRPLG
DSNGCPCGARRVDAPDLYGSGFLPPCPVCGRVAVGPTVREACGCHVTNCACETACA
VORLCIGGGRRLVAILCDYPPYGPWRGPRTRPEAHEAVQRSSESDACTCAP"
129384..129403
129446..129451
/note="ORF" s 8 and 9"
129746..130201
/note="potential product; ORF 10; Identical sequence as
found in span 14076-14531; potential membrane protein;
related sequence to ORF 9, ORF 11; putative; NCBI gi:
331297"
/codon_start=1
/number=10
/product="membrane protein"
/translatoins="MTVRGCLCLAFGVTLIVGVVYVAVGSKGRAPGDDLVTLAL
NELAGDEWVGRISDAGANITGLRGLILDLHSRNSGSLVADALYEVILTRLRLDI
FYFNGVAAPNGWFLSFIIMTCAITVRSKKVAPAPHSPTTMSDPI"
130209..131054
/note="potential product; ORF 11; Identical sequence as
found in span 14539-15384; potential Zn-binding protein;
related sequence to ORF 9, ORF 10; putative; NCBI gi:
331298"
/codon_start=1
/number=11
/product="Zn-binding protein"
/translation="MRCIRAPADDDGYTPIIMSPVPCFCQCEAVFVSNRASCHV
FCFFCTRECECGSGGRKLIPNEYLVATKAPFPAPMGRTAGTWMGPNQGWHE
PRVWLEDLITAVITVGAIVETRGACGVAVRARKWEGAIALPLDLLDLVELGCG
AIEACGKVAAGCVLVRTLEIVLVRYVHDTLAKTFPVMAFRFCSLGALKELLSRFRIPG
YFGSGTRYDGLLGHIAICEIRKCCDDPGIGFPNPFPAIRNRKNHKKRTGGATGSPFTD
PTGH"
131061..131960
/note="potential product; ORF 12; Identical sequence as
found in span 15391-16290; potential Zn-binding protein;
putative; NCBI gi: 331299"
/codon_start=1
/number=12
/product="Zn-binding protein"
/translation="WTSPREIIVRLSCGHPLRVPDGEFCFGDCAVCCDCDERAVFFA
PHLHTVLSVAPGDLGALTEKSMILAKAIIVTGCDDVLQYPIVHERGHWGCRFCF
RQKCAEGCLMARTVAPLDLIERLQESLFGHTRGLEYSKDFETAEIWMGPOGKEVILYI

```

		DGALGVNFVDVDTVGISNVIADRAKEP IDLDIVILANPDWSPVPCTIPYRLERLIIWKANF ESGWHDMSDLDTLGGANNIIVGAVFGDGRSLILEFLIRYLIVDPGEAFETPAUTFEFF
		LISLAECTWVGDSRGCGYDRCSR"
polyA_signal		131999..132004
		/note="ORF" # 10, 11 and 12"
CDS		132024..132272
		/note="ORF" 13; Identical sequence as found in span
		16354-16602; NCBI gi: 331300"
		/codon_start=1
		/number=13
		/translation="MAFLIFLFICNCNPMSLLCCGGGDLISSCCRGCGQWPMPARQPIY PYSGPMGAHYVPPVAQPVRGPRVPOGERPVDELGR"
repeat_unit		132335..132386
polyA_signal		132410..132415
		/note="ORF" 13"
polyA_signal		132414..132419
		/note="ORF" 13"
repeat_unit		132717..133026
polyA_signal		complement(132746..132751)
		/note="ORF" 14"
polyA_signal		complement(132798..132803)
		/note="ORF" 14"
polyA_signal		complement(132849..132854)
		/note="ORF" 14"
polyA_signal		complement(132901..132906)
		/note="ORF" 14"
polyA_signal		complement(132968..132973)
		/note="ORF" 14"
CDS		complement(133016..134191)
		/note="ORF" 14; Identical sequence as found in span
		17346-18521; related sequence to ORF 15, ORF 16; NCBI gi: 331301"
		/codon_start=1
		/number=14
		/product="protein Kinase"
		/translation="MAEKLIPTSCGFPLPARVLREGLRAPRVADVDDGFKLTIGPLT SYLRMDVYHYNENATDWLRGNPFWEEDCDVIDLPGVACVNEATFLTKKKYTG VTLHFHFNLRHTNRITFFGSP LSAATGMALIKTFHGQTQAPDRPGMEYIAEMTA FLDAAPKLIIEWEYSGTRENKSAOWTFSEGLYSGMEGYFERATVDICATIALD LTEKLALARKGITVHGDLSKENIIMSRSGPKGLIDFEHSHPGETWTSFWTPDRFTF WNPTGEAYAPERSRGRNGRPVGIVFQGLIALINIMVERMERVFVNHTWIKGDGY RAHVLYKVKARTLIDLRRGARTLARVDIELIGVARCLERDAMPRSLETLVDEFSKI" BASE COUNT 28727 a 37766 c 37707 g 30026 t
	ORIGIN	
	DB 71; Score	24; Match 78.6%; QryMatch 1.6%; Pred. No. 1.46e-01;
	Matches	33; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Db	42146	ggaccttcgacagaagctcgaaagtcaacttaccacgagc 42187
Cp	679	GGACCTTCACCAAGACGAGCAGGCACCTTTCCCAGAAGC 638
RESULT	11	
LOCUS	TRECRI1	3940 bp DNA INV 15-SEP-1990
DEFINITION	C.fasciculata retrotransposable element I (CRE 1) gene complete cds.	
ACCESSION	M33009	
KEYWORDS	integrase; retrotransposable element I; reverse transcriptase.	
SOURCE	C.fasciculata.	
ORGANISM	Crithidia fasciculata	
	Eukaryota; Animalia; Protozoa; Sarcosastigophora; Mastigophora; Kinetoplastida; Trypanosomatina; Trypanosomatidae.	
REFERENCE	1 (bases 1 to 3940)	

AUTHORS Gabriel, A., Yen, T.J., Schwartz, D.C., Smith, C.L., Boeke, J.D.,
Sollner-Webb, B. and Cleveland, D.W.
TITLE Rapidly rearranging retrotransposon within the minixon gene locus
of *Crithidia fasciculata*
JOURNAL Mol. Cell. Biol. 10, 615-624 (1990)
MEDLINE 90136575
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by A.Gabriel, 21-MAR-1990.

FEATURES NCBI gi: 162492
source Location/Qualifiers
1..3940
/organism="Crithidia fasciculata"
repeat_region 1..415
/note="mini-exon gene repeat"
misc_feature 386..415
/note="target site duplication"
misc_feature 415
/note="mini-exon splice junction"
misc_feature 3899..3927
/note="target site duplication"
repeat_region 3899..3940
/note="mini-exon gene repeat"
BASE COUNT 890 a 1032 c 1330 g 688 t
ORIGIN

DB 31; Score 24; Match 67.6%; QryMatch 1.6%; Pred. No. 1.46e-01;
Matches 46; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Db 2493 atggtgtgtggccgaagcaaccgggtggggtgacgtggggtgacgtgggtgtg 2552
|||||
|||||

Cp 1174 ATGCTGGTGGAGGCTCAGGAACAGGGGTGGAGCTGGGTGGATCCATGGGTGTG 1115

Db 2553 gccgagggc 2560
|||||
Cp 1114 GGCAGGC 1107

RESULT 12
LOCUS MUS05F1 1514 bp mRNA ROD 13-JAN-1992
DEFINITION Mouse mRNA for OSF-1.
ACCESSION D90225
KEYWORDS OSF-1.
SOURCE Mouse (new born; strain C57BL/6) calvaria osteoblast cell, cell
line MC3T3-E1, cDNA to mRNA, clone pMCO31.

ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1514)
AUTHORS Eukaryota; Animalia; Chordata; Mammalia; Rodentia; Muridae.
Tezuka, K., Takeshita, S., Hakeda, Y., Kumegawa, M., Kikuno, R. and
Hashimoto-G, T.
TITLE Isolation of mouse and human cDNA clones encoding a protein
expressed specifically in osteoblast and brain tissues
JOURNAL Biochem. Biophys. Res. Commun. 173, 246-251 (1990)
MEDLINE 91076859
COMMENT These data kindly submitted in computer readable form by: Reiko
Kikuno
Pharma Research Labs.
Hoechst Japan Ltd.
1-3-2 Minamidai
Kawagoe, Saitama 350
Japan
Phone: 492-43-6149
Fax: 492-45-6036.

FEATURES NCBI gi: 220552
source Location/Qualifiers
1..1514
/organism="Mus musculus"
CDS 215..721
/gene="osf-1"
/note="NCBI gi: 220553"
/codon_start=1
/product="OSF-1"
/translation="MSSQQYQQRRKFAAFILAFILAAVDTAEGKKEPKKVKK
SDCGEMQWVCVPTSGDGLGTREGTRTGAECQTKMTQRCIPCNWKKQFGAECYKQ
FOAGCEDLNTALKTRTGSIKRALHWDQRTVTIISKPCGKLTGPKQAESKKKKEG
KKQEKMLD"

polyA_signal 1476..1481
BASE COUNT 484 a 267 c 350 g 413 t
ORIGIN

DB 57; Score 24; Match 71.4%; QryMatch 1.6%; Pred. No. 1.46e-01;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 135 cctggagctgcgaacactgcgcaccttgcacaagaagcagcagctagtca 190
|||||

Qy 980 COTGAGGCTCAGACATCTGCCCGGAGCTTTCACAAAGGCTCCCTGGCATCA 1035

RESULT 13
LOCUS HUMPRD1GEN 53522 bp DNA PRI 28-APR-1995
DEFINITION Homo sapiens polycystic kidney disease-associated protein (PKD1)
gene, complete cds.

ACCESSION L39891
KEYWORDS polycystic kidney disease.
SOURCE Homo sapiens (clone: cg9g10.2, cDEB11) DNA.
ORGANISM Homo sapiens

Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Homnidae.
REFERENCE 1 (sites)
AUTHORS Ward, C.J., Peral, B., Hughes, J., Thomas, S., Gamble, V.,
MacCarthy, A.B., Sloane-Stanley, J., Buckle, V.J., Kearney, L.,
Higgs, D.R., Ratcliffe, P.J., Harris, P.C., Roelfsema, J.H.,
Spruit, L.L., Sarris, J.J., Dauwerse, H.G., Peters, D.J.M.,
Breuning, M.H., Nellist, M., Brook-Carter, P.T., Maheshwar, M.M.,
Cordeiro, I., Santos, H., Cabral, P., Sampson, J.R., Janssen, B.,
Hesselink-Jansen, A.L.W., van den Ouweland, A.M.W., Eussen, B.,
Verhoeve, S., Lindhout, D. and Halley, D.J.J.

The polycystic kidney disease 1 gene encodes a 14 kb transcript and
lies within a duplicated region on chromosome 16. The European
Polycystic Kidney Disease Consortium [published erratum appears in
Cell 1994 Aug 26;78(4):following 724]

JOURNAL Cell 77 (6), 881-894 (1994)
MEDLINE 94273192
REFERENCE 2 (bases 1 to 53522)

AUTHORS Burn, T.C., Connors, T.D., Dackowski, W.R., Petry, L.R., Van Raay, T.J.,
Milholland, J., Venet, M., Miller, G., Hakim, R.M., Doggett, N.A.,
Landes, G.M., Klinger, K.W., Qian, F., Onuchic, L.F., Watnick, T. and
Germino, G.G.

TITLE Analysis of the genomic sequence for the autosomal dominant
polycystic kidney disease (PKD1) gene predicts the presence of a
leucine-rich repeat

JOURNAL Hum. Mol. Genet. 4 (4), 575-582 (1995)
COMMENT NCBI gi: 790818

FEATURES Location/Qualifiers
source 1..53522
/organism="Homo sapiens"
/clone="cg9g10.2, cDEB11"

```

/resequenced_mol="DNA"
complement(236..524)
/rpt_family="Alu"
complement(704..940)
/rpt_family="Alu"
complement(998..1287)
/rpt_family="Alu"
complement(998..1287)
/rpt_family="Alu"
complement(1483..1771)
/rpt_family="Alu"
complement(2081..2352)
/rpt_family="Alu"
complement(2423..2666)
/rpt_family="Alu"
3096..4204
/notes="CpG island; putative"
<3648..3862
/genes="PKD1"
/map="16p13.3"
/notes="putative"
/number=1
join(3648..3862,15765..15817,19866..19974,20096..20167,
20436..20605,21110..21490,21609..21792,22228..22448,
22637..22752,23118..23289,23656..23815,24967..25111,
25989..26120,26318..26493,26808..26941,27407..27525,
27628..28683,29126..29875,30419..31024,31244..31393,
32328..32972,33185,33251..33410,
33848..33952,36954..37098,37701..38330,38624..38780,
38961..39213,39338..39533,41028..41198,41285..41428,
41523..41733,41824..41950,43610..43726,43817..43866,
44091..44275,44353..44446,47384..47502,47581..47783,
47856..48050,48501..48640,49002..49114,49406..49547,
49688..49813,49997..50171,50419..50709,50785..50919,
51003..51308,51399..51866)
/genes="PKD1"
/map="16p13.3"
/notes="NCBI gi: 790819"
/codon_start=1
/product="polycystic kidney disease-associated protein"
/translation="MPPAAPRLAALGLGLMGLAGGGRGCGCEPPCLCGPAPG
AACRVNCSGRGLRTGLPALRIPADATALTAMAKADGVKKYKSGKGAALATHGSLILR
DYSNHLRALDVLGLANLSALAEIDISNNKISTLEEIPANLENLSGNPFEC
GGLAMLPRAEEQQRVVPQPEAATCAGPGSLAGQPLILGIPLLDSCGATLVGPGPLA
SGOLAAPHIAAPLPVTAATRWDFGCGSAEVDAAAPAAASHRYLPGRVHTVAIAGCS
ALLCTDVQVEAPAALELVCPSVQSQSDSIDLSIQNRGSGLEAAYSIVALGEEPARA
VHP LCP SDTEIFPGNGHCIRIVVEKAAMQAQEQCQWAGAAALMVDSPAVQRFIVSR
VTRSLDWITGFTVQVGVGAPQCEAFSIESCQWLPGEPPATAEHCVRCLGTCG
NTDLCSPHUSVYCELPQCGGVQDAENLLCAPSGDLOGPLTQAQDGLSAPHEVEA
GNGVELRPPAHQVWVFGELISREAF LTTAEF GTQLRRPAQLRLQVTRLSTAGT
PENGSEFDRTPQTLQACMFGRMGFGANTCLPLDASCHPQACANGCTEGEHV
DVVENSASRANLSIRVTAEEP IGLBRATPSPEARVLQGVIVRVSPVVEAGSDVFM
TINDQSLFTQNVFNVIQSAAVFKLSLTASNHVSNTVNTVNTVEMNRMQGLQVS
TVPLVLSPLATLALTCGLVDAVEAVFLWTFCDGEQALHQFPQPNESFPYDPQSV
QVIVHNVMHTYAPGEYLLTVLASNAFENRTQVVPVSRSLPSVAVGSDVGLSP
RANHVTAGTTHYVRLEVNNTVSGAAQADRVFEELRGLSVDASLAVEGAPVWSA
AVQTGDI NTDFMGDCGTVLSGPETAHVHYLRAQNCITVGAASPGAHLSRLVIL
FVLEVRPEPAACITPTOPARLTAVYTGPNPAHYLFDWTFDGGSNNTVRCPTVTHNF
TRSGTTP LALVLSRVRNRAHYETSICVEPEVGNVTLQPERQVQLGDEAWLVACWPP
FFYRYTDFGTEEAAPFRARGPEVTFIYRDPGSYLVTVTASNISAANDSALVEQPP
VLWTSIKWNGSLGLELQAPYLP SAGRCRPA SYLWLDGCGMLEGPEVTHAYNSTGDF
TGP GFGQRQRLLAHLEAGTYHYQLRATNMLGSAWADCTHDFVEPVCHLWASPNP
AAVNKSVTLAELAGSGVGYTW SLEEGSWETSEPTTTHSPTTTPGLHVTWTAGNPL
GSANATVEVDVQVPVSGLSIRASEPGSGFVAAGSVFPFWQLATCTGNSMCNAVPGS

```

```

SKGPHVTWFPDAGTF S IRLNASHNAVSWSATNTLTAEEP IGLVLMASKRVAPGQ
LVHFQILLAGSAGTFR LQDTPDEAEHS YLRPGDYRVQVNASNLVSFFVAQATVTVQ
VLACREPDVWLP LQVLMRESORNYLEAHVDLRDCTVYOTEYRWEVYRTASCQRCR
PARVALPGVDVSRRLV LRLALPVGHYCFVVFVSGDTP LTOS IQANTVVAPELVP
IIEGGSYRVSDTRDLVDGSESDPNLDGQTP LSFHWACVASTOREAGCCALNF
PGSSTVTI PRERLAAGVEYFSLTVKAGRKEATNQTILRS GRVP IVSLECVGCK
AQAVEYSRSYVY LEGRCLNCSSCKRGAARTFSNKT LIVLDETTTSTGSAOMRLV
LRRGLRDGEGYTF TLTVLGRSEEGGAS IRLSPNRPLGGSCRLFP LGAVHALTK
VHECTGWHDAEDAGAPLVALLRRQCCHCEFCVKGSLSGYAGVLPFGFRHFE
VGLAVVQDQLGAAVVALNRSLATLPEPGSATGLTVLHGLTASVLPGLLRQADPQ
HVTEYSALAVTVINEHRAQIRKNTETIVLSIRVHTVDDI IQIATAALACMQCPSRELVC
RSCIKQTLHKLEAMLI LQAETAGTVPPTAIGDSITANTGDLIHLASSDVAPQFSE
LGAESP SRMASOAYNLTSALMRI LMRSRVLINEEP LTAGEEIVAOGRSDPRLSLCY
GGAPG GCHFSIPEAFSGALANLSDVQLIFLVDSPNPPFGYISNTVSTKVASMAFQ
TSNPAAGLHQJLNTYLLDGHVYLSEEPYLAUVLHSEPRPNEHNCASRRIRPESLQG
ADHRPYTFE IPSGRDPAGSYHLNLSHFWSALQVSVGLYLSLCQYFSEEDWWRTE
GLIPLSETPSRQAVCLTRHLTAF GASLFVPPSHVREVFPEPTADVNY IVMLTCAVCLV
TYMMAATLHKLDQ LIDASGRAIPFCQGRGRFYELLVKTKGRCGSGTTAHVGLMYG
VDSRSGHRHLDGDRAFHNSLDIFRIATPHSLGSMKIRVWHDKGLSPAMFLQHVIV
RDIQTARSAFFLVNMDLVSYTEANGCLVEKEVIAAADAALIRFRLLVLAELQGRFDDK
HIMLSITWRPPRSFRTRQIATCCVILLICIFLGRMSKRVAGSPPTPAQQVLDIDSLDSSVL
SDTVAVAGLSVSVVVPYVYTAI LFLFRMSKRVAGSPPTPAQQVLDIDSLDSSVL
DSSF LTFSGLHAEAFVGMKSDLF LDDSKSLVCMPSGEGT LSWPDLSDPSIVGSNLR
QIARQAQHGIGLPEDEGFSLASPYSPAKSF SASDEDLIQVLAEGVSPAPTQDTHME
TDLSSLSSTPGKTELTALQRLGELGPPSPGLNWEQPOAARLSRTGLVEGRKRLIP
AWCASLHGLSLLLVAVAVGSGVAGSPFPVGVSVAMLLSSSASF LASF LQWEP LKVL
LEALYFSLVAKRLHDEDDTLVESPAVTVSARVPRVRPPHGFALFLAKEARKVKRL
HGMRLSLVYMLFLVITLLASYGDSACHAYRLQSAIKOELHSRAFLAITRSEELWP
WMAHVLPLYVHGNQSGFELGPPRLVRQLRQALYLPDPPGPRVHTCSAAGSTSDVDV
WGNRSHNGSGTWAY SAPDILLCAMSGSCAVDYSGGYVQELGSLSEESRDLRFLQLHN
WLNORSRAVLELTRYSPAVGLHAAYTLRLLEFFAAGRALAALSVRPFALRLRSLAGLSL
PLITSVCLLLEFANVFAVEAETHWREGRWVRLIGAWARLLVALTAATALVRLAQIG
AADORTFVRGRPRRTFSDQVAQSSRWARGLAASLIF LLLVRAAQQLRFVROMSVF
GKTLCHRALPELLGVTGLVGVAYQA LAILVSSCVDLSMSVAQALIVLCPGTGLST
LCPAESWHLSP LLLCVGLMALRLMCA LRGAVILRMRYHALRGELYRPAWEPQYEMVE
LF LRLRLMGLSKVKEFRHKVFCMEQPLPSRSRSGKSVDPVPPSAGSDASHST
SSQLDGLSVSLGRLGTCEPEP SRIQAVFEALTQFDRLNQAATEDVYQLQQLHSIQ
GRSSRAPAGSSRGPSPGLRPLPSRLASRQVLDLATCPSRTP LRANKVHPSSI"
3863..15764
/genes="PKD1"
/map="16p13.3"
/number=1
complement(6251..6540)
/rpt_family="Alu"
complement(6758..7024)
/rpt_family="Alu"
complement(7082..7371)
/rpt_family="Alu"
complement(7510..7763)
/rpt_family="Alu"
complement(7821..8107)
/rpt_family="Alu"
9490..9774
/rpt_family="Alu"
10178..10411
/rpt_family="Alu"
11958..12187
/notes="CpG island; putative"
15765..15817
/genes="PKD1"
/map="16p13.3"
/notes="putative"
/number=2

```

intron

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

misc_feature

exon

```

intron      15818..19865
            /gene="PKD1"
            /map="16p13.3"
            /number=2
repeat_region complement(17907..18198)
repeat_region /rpt_family="Alu"
repeat_region complement(18260..18516)
repeat_region /rpt_family="Alu"
repeat_region complement(18574..18843)
repeat_region /rpt_family="Alu"
repeat_region complement(19064..19350)
repeat_region /rpt_family="Alu"
exon        19866..19974
            /gene="PKD1"
            /map="16p13.3"
            /number=3
intron      19975..20095
            /gene="PKD1"
            /map="16p13.3"
            /number=3
exon        20096..20167
            /gene="PKD1"
            /map="16p13.3"
            /number=4
intron      20168..20435
            /gene="PKD1"
            /map="16p13.3"
            /number=4
exon        20436..20605
            /gene="PKD1"
            /map="16p13.3"
            /number=5
intron      20606..21109
            /gene="PKD1"
            /map="16p13.3"
            /number=5
exon        21110..21490
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=6
misc_feature 21174..21459
            /gene="PKD1"
            /map="16p13.3"
            /note="CpG island; putative"
intron      21491..21608
            /gene="PKD1"
            /map="16p13.3"
            /number=6
exon        21609..21792
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=7
intron      21793..22227
            /gene="PKD1"
            /map="16p13.3"
            /number=7
exon        22228..22448
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=8
intron      22449..22636

```

```

exon        22637..22752
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=9
intron      22753..23117
            /gene="PKD1"
            /map="16p13.3"
            /number=9
exon        23118..23289
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=10
intron      23290..23655
            /gene="PKD1"
            /map="16p13.3"
            /number=10
exon        23656..23815
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=11
intron      23816..24966
            /gene="PKD1"
            /map="16p13.3"
            /number=11
exon        24967..25111
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=12
intron      25112..25988
            /gene="PKD1"
            /map="16p13.3"
            /number=12
exon        25989..26120
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=13
intron      26121..26317
            /gene="PKD1"
            /map="16p13.3"
            /number=13
exon        26318..26493
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=14
intron      26494..26807
            /gene="PKD1"
            /map="16p13.3"
            /number=14
exon        26808..26941
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=15
intron      26942..27406
            /gene="PKD1"

```

	/map="16p13.3"
	/number=15
misc_feature	27407..28254
	/note="CpG island; putative"
exon	27407..27525
	/gene="PKD1"
	/map="16p13.3"
	/note="putative"
	/number=16
intron	27526..27627
	/gene="PKD1"
	/map="16p13.3"
	/number=16
exon	27628..28683
	/gene="PKD1"
	/map="16p13.3"
	/note="putative"
	/number=17
misc_feature	28608..28993
	/note="CpG island; putative"
intron	28684..29125
	/gene="PKD1"
	/map="16p13.3"
	/number=17
exon	29126..29875
	/gene="PKD1"
	/map="16p13.3"
	/note="putative"
	/number=18
misc_feature	29857..30825
	/note="CpG island; putative"
intron	29876..30418
	/gene="PKD1"
	/map="16p13.3"
	/number=18
exon	30419..31024
	/gene="PKD1"
	/map="16p13.3"
	/note="putative"
	/number=19
intron	31025..31243
	/gene="PKD1"
	/map="16p13.3"
	/number=19
exon	31244..31393
	/gene="PKD1"
	/map="16p13.3"
	/note="putative"
	/number=20
intron	31394..32327
	/gene="PKD1"
	/map="16p13.3"
	/number=20
exon	32328..32471
	/gene="PKD1"
	/map="16p13.3"
	/note="putative"
	/number=21
intron	32472..32598
	/gene="PKD1"
	/map="16p13.3"
	/number=21
exon	32599..32878
	/gene="PKD1"

misc_feature	/map="16p13.3"
	/note="putative"
intron	/number=22
	32616..33079
exon	/note="CpG island; putative"
	32879..32971
intron	/gene="PKD1"
	/map="16p13.3"
exon	/number=22
	32972..33185
intron	/gene="PKD1"
	/map="16p13.3"
exon	/note="putative"
	33186..33250
intron	/number=23
	/gene="PKD1"
exon	/map="16p13.3"
	33251..33410
intron	/gene="PKD1"
	/map="16p13.3"
exon	/note="putative"
	33411..33847
intron	/gene="PKD1"
	/map="16p13.3"
exon	/number=24
	33848..33952
intron	/gene="PKD1"
	/map="16p13.3"
exon	/note="putative"
	33953..36953
intron	/gene="PKD1"
	/map="16p13.3"
exon	/number=25
	36954..37098
misc_feature	/gene="PKD1"
	/map="16p13.3"
intron	/note="putative"
	37007..37251
exon	/note="CpG island; putative"
	37099..37700
intron	/gene="PKD1"
	/map="16p13.3"
exon	/number=26
	37701..38330
intron	/gene="PKD1"
	/map="16p13.3"
exon	/note="putative"
	38331..38623
intron	/gene="PKD1"
	/map="16p13.3"
exon	/number=27
	38624..38780
intron	/gene="PKD1"
	/map="16p13.3"
exon	/number=28
	38781..38960
intron	/gene="PKD1"
	/map="16p13.3"
exon	/number=28
	38961..39120

[illegible][illegible]

```

exon      38961..39213
          /gene="PKD1"
          /map="16p13.3"
          /number=29
intron    39214..39337
          /gene="PKD1"
          /map="16p13.3"
          /number=29
exon      39338..39533
          /gene="PKD1"
          /map="16p13.3"
          /number=30
intron    39534..41027
          /gene="PKD1"
          /map="16p13.3"
          /number=30
repeat_region complement(39970..40228)
          /rpt_family="Alu"
repeat_region 40311..40598
          /rpt_family="Alu"
exon      41028..41198
          /gene="PKD1"
          /map="16p13.3"
          /number=31
intron    41199..41284
          /gene="PKD1"
          /map="16p13.3"
          /number=31
exon      41285..41428
          /gene="PKD1"
          /map="16p13.3"
          /number=32
intron    41429..41522
          /gene="PKD1"
          /map="16p13.3"
          /number=32
misc_feature 41501..41730
          /note="CpG island; putative"
exon      41523..41733
          /gene="PKD1"
          /map="16p13.3"
          /number=33
intron    41734..41823
          /gene="PKD1"
          /map="16p13.3"
          /number=33
exon      41824..41950
          /gene="PKD1"
          /map="16p13.3"
          /number=34
intron    41951..43609
          /gene="PKD1"
          /map="16p13.3"
          /number=34
repeat_region 42463..42685
          /rpt_family="Alu"
repeat_region 43084..43321
          /rpt_family="Alu"
exon      43610..43726
          /gene="PKD1"
          /map="16p13.3"
          /number=35
intron    43727..43816
          /gene="PKD1"

```

```

          /map="16p13.3"
          /number=35
exon      43817..43866
          /gene="PKD1"
          /map="16p13.3"
          /number=36
intron    43867..44090
          /gene="PKD1"
          /map="16p13.3"
          /number=36
exon      44091..44275
          /gene="PKD1"
          /map="16p13.3"
          /number=37
intron    44276..44352
          /gene="PKD1"
          /map="16p13.3"
          /number=37
exon      44353..44446
          /gene="PKD1"
          /map="16p13.3"
          /number=38
intron    44447..47383
          /gene="PKD1"
          /map="16p13.3"
          /number=38
repeat_region 44778..45002
          /rpt_family="Alu"
repeat_region complement(46331..46602)
          /rpt_family="Alu"
exon      47384..47502
          /gene="PKD1"
          /map="16p13.3"
          /number=39
intron    47503..47580
          /gene="PKD1"
          /map="16p13.3"
          /number=39
exon      47581..47783
          /gene="PKD1"
          /map="16p13.3"
          /number=40
intron    47784..47855
          /gene="PKD1"
          /map="16p13.3"
          /number=40
exon      47856..48050
          /gene="PKD1"
          /map="16p13.3"
          /number=41
intron    48051..48500
          /gene="PKD1"
          /map="16p13.3"
          /number=41
exon      48501..48640
          /gene="PKD1"
          /map="16p13.3"
          /number=42
intron    48641..49001
          /gene="PKD1"
          /map="16p13.3"
          /number=42
exon      49002..49114
          /gene="PKD1"

```


Db 94 cggtcagacaaaggcgagtgggcccaggctgggccctggaactggaggagtg 153
||| ||| ||| | |||| | |||| | |||| ||||
Qv 400 CGGACAGAGCAGGCACAGCACATTCAGGGGCAGTCCCTTCTACTGGAGGGAGTGA 459

Db 154 agg 156
Ov 460 TGG 462

RESULT	15			
LOCUS		CHKOVAL5A	5322 bp	DNA
DEFINITION		Chicken ovalbumin gene, 5' flank.		
				VRT
				15-MAR-1990

ACCESSION M29020
KEYWORDS ovalbumin.
SOURCE Chicken DNA, clones pBR EcoRI(6,7)
ORGANISM Gallus gallus

Eukaryota; Animalia; Chordata; Vertebrata; Aves; Neornithes;
Neognathae; Galliformes; Phasianidae.
1 (bases 1 to 532)

AUTHORS Kaye, J.S.
JOURNAL Unpublished (1989) U. Rochester, Dept. Biol., Rochester, NY 14627
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted by J.S.Kaye. 14-OCT-1989.

```

NCBI gi: 212507
FEATURES
    source          Location/Qualifiers
                    1..5322
                        /organism="Callus callus"

```

BASE COUNT	ORIGIN
1469 a	1240 c 1103 g 1510 t

DB 73; Score 23; Match 87.1%; QryMatch 1.5%; Pred. No. 7.59e-01; Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 3442 gtactgcttgacagctgtgctgggcagg 3472
 ||||| | ||||| | |||||
 Cp 730 GTACTGCTTGGGACAGCTGTGGTGGCAGG 700

Search completed: Thu Sep 28 00:03:27 1995
Job time : 844 secs.

22 21 1.4 302 32 RIC07191A Rice cDNA, partial se 1.13e-02
 23 21 1.4 366 19 R16830 yf34a12.s1 Homo sapie 1.13e-02
 24 21 1.4 558 23 R31294 yb60q12.s1 Homo sapie 1.13e-02
 25 21 1.4 269 31 R59677 yb02q01.r1 Homo sapie 1.13e-02
 26 21 1.4 468 24 R35940 yb68e06.r1 Homo sapie 1.13e-02
 27 21 1.4 375 32 R63283 y107f04.r1 Homo sapie 1.13e-02
 28 21 1.4 472 27 R46605 yb51h10.s1 Homo sapie 1.13e-02
 29 21 1.4 328 23 R32306 yb68b02.s1 Homo sapie 1.13e-02
 30 21 1.4 352 52 T77909 yd16b02.r1 Homo sapie 1.13e-02
 31 20 1.3 258 58 T97901 ye54b12.r1 Homo sapie 1.48e-01
 32 20 1.3 312 58 T98862 ye65b10.s1 Homo sapie 1.48e-01
 33 20 1.3 432 55 T87706 yd91h12.s1 Homo sapie 1.48e-01
 34 20 1.3 278 9 HSC2HC102 H. sapiens partial cd 1.48e-01
 35 20 1.3 204 57 T93428 lmeST0196 Leishmania 1.48e-01
 36 20 1.3 411 39 T27238 HCOEST58 Homo sapiens 1.48e-01
 37 20 1.3 436 16 R05578 yf87g06.r1 Homo sapie 1.48e-01
 38 20 1.3 457 25 R38972 yd07d08.s1 Homo sapie 1.48e-01
 39 20 1.3 475 29 R52272 yj73ell.r1 Homo sapie 1.48e-01
 40 20 1.3 485 33 R1GS12189A C.elegans cDNA clone 1.48e-01
 41 20 1.3 355 2 CELK01607R Rice cDNA, partial se 1.48e-01
 42 20 1.3 390 54 T83445 yd46c07.r1 Homo sapie 1.48e-01
 43 20 1.3 539 53 T79953 yd85c11.s1 Homo sapie 1.48e-01
 44 20 1.3 411 33 RICR3023A Rice cDNA, partial se 1.48e-01
 45 20 1.3 202 41 T33080 EST56609 Homo sapiens 1.48e-01

ALIGNMENTS

RESULT 1
 LOCUS T16981 398 bp mRNA EST 01-AUG-1994
 DEFINITION NB2099-5R Homo sapiens cDNA 5' end.
 ACCESSION T16981
 KEYWORDS EST.
 SOURCE human library=Normalized infant brain, Bento Soares vector=BA, M13-derived host=E. coli DH5-alpha primer=M13 Reverse
 Reitel=HindIII Reitel2=NotI The normalized infant brain library, constructed by Bento Soares, Columbia University, was oligo-(dT) primed and directionally cloned into an M13-derived plasmid using total brain mRNA from a 72-day old human female afflicted with spinal muscular atrophy. The library was normalized as described elsewhere.

ORGANISM Homo sapiens
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 398)
 AUTHORS Stevens,T.J., Berry,R., Goold,R., Walter,N.A.R., Wilcox,A.S., Hopkins,J.A., Rubano,T., Weber,J., Soares,M.B. and Sikela,J.M.
 TITLE Gene-based STSS as the basis for a human gene map
 JOURNAL Unpublished (1994)
 COMMENT Contact: Sikela JM
 Department of Pharmacology
 University of Colorado Health Sciences Center
 Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
 Tel: 3032706884
 Fax: 3032707097
 Email: tjgaetally.hsc.colorado.edu.

NCBI gi: 519143
 Location/Qualifiers
 source 1..398
 /organism="Homo sapiens"
 /note="human"

BASE COUNT 100 a 117 c 75 g 101 t 5 others

ORIGIN
 DB 37; Score 25; Match 62.1%; OryMatch 1.7%; Pred. No. 1.10e-07;
 Matches 64; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
 Db 70 cctctctctctgagttgaccagcagggtctgcccagccaccagcaccatctctctc 129
 ||||| ||| ||| ||| ||| | | | | | ||||| ||| ||| ||| ||| ||| |||
 Qy 1095 CTTCTTCACTGGCTTGGCCACCACCCATGATCTCCACCCAGCTCCACCCCTGTTT 1154
 ||||| ||| ||| ||| ||| | | | | | ||||| ||| ||| ||| ||| ||| |||
 Db 130 ccacgagctccagaaccatgccaggtctctctctcacatc 172
 || | |||| ||||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Qy 1155 CTTGACCTTCACCATGCTTAACCTACCGCCCTCATC 1197

RESULT 2
 LOCUS R17453 436 bp mRNA EST 14-APR-1995
 DEFINITION y914d02.r1 Homo sapiens cDNA clone 32427 5'.
 ACCESSION R17453
 KEYWORDS EST.
 SOURCE human clones=32427 library=Soares infant brain INIB vector=Lafmid BA host=DH10B (ampicillin resistant) primer=M13Rpl Reitel=Not I Reitel2=Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAGAAATTCGCGCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 436)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Ruffin,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Riffkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT GDB: G00-404-774
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 High quality sequence stops: 399
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

NCBI gi: 771063
 Location/Qualifiers
 source 1..436
 /organism="Homo sapiens"
 /clone="32427"
 /note="human"

BASE COUNT 84 a 128 c 121 g 100 t 3 others
 ORIGIN
 DB 19; Score 23; Match 69.5%; OryMatch 1.5%; Pred. No. 4.44e-05;

Matches	41;	Conservative	0;	Mismatches	18;	Indels	0;	Gaps	0;
Db	246	tctgtgtcgtcttccccaccagcagtgccctccccatccccctctctgtttc	304						
Qy	1097	TCTTTCACCTGGCTGGCCACCACCACTGGATCTCACCCGAGCTCCACCCCTGTTTC	1155						
RESULT	3								
LOCUS		HSCIWE031	368 bp	RNA	EST			06-NOV-1994	
DEFINITION		H. sapiens partial cDNA sequence; clone c-lwe03.							
ACCESSION		Z44264							
KEYWORDS		partial cDNA sequence; transcribed sequence fragment.							
SOURCE		human.							
ORGANISM		Homo sapiens							
		Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chondata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE		1 (bases 1 to 368)							
AUTHORS		Genexpress.							
TITLE		Direct Submission							
JOURNAL		Submitted (24-OCT-1994) to the EMBL/GenBank/DDBJ databases.							
		Genethon, B.P. 60, 91002 Evry Cedex France and Genetique Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr							
REFERENCE		2 (bases 1 to 368)							
AUTHORS		Genexpress.							
TITLE		The Genexpress cDNA program							
JOURNAL		Unpublished							
REFERENCE		3 (bases 1 to 368)							
AUTHORS		Auffray, C., Behar, G., Bois, F., Boucher, C., da Silva, C., Devignes, M.D., Duprat, S., Houllatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani-Kabatchis, C. and Tessier, A.							
TITLE		IMAGE: Integrated molecular analysis of the human genome and its expression							
JOURNAL		C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)							
COMMENT		Clone library from B.Soaress, Psychiatriy Dept. Columbia University USA;							
		Cloning method: total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector;							
		Sequencing method: single read, full automatic;							
		Primer: M13 reverse							
		cDNA sequence colinear to mRNA							
		Stretch removed: nothing							
		Normalization method: Bento Soares, P.N.A.S in press;							
		Genexpress_library_idt: C;							
		Genexpress_sequence_idt: y1c-lwe03;							
		No significant homology found with :							
		genbank release 81 swissprot release 28.							
FEATURES		NCBI gi: 573389							
source		Location/Qualifiers							
		1..368							
		/organism="Homo sapiens"							
		/dev_stage="3 months old"							
		/isolate="muscular atrophy patient"							
		/tissue type="total brain"							
		/clone_lib="normalized infant brain cDNA"							
		/sex="Female"							

Cp	1384	TCTCAGCTGCTGTATTCCTTCCCGAGGGCGCTTT	1349
II IIIIIIIII III IIIIIII IIIIIIIIIII			
-	RESULT	5	
	LOCUS	R35295	236 bp mRNA
	DEFINITION	Y662409.r1 Homo sapiens cDNA clone 37553 5'	EST 02-MAY-1995
	ACCESSION	R35295	
	KEYWORDS	EST.	
	SOURCE	human clones-37553 library-Soares infant brain IN1B vector-Lafmid BA host-DH10B (ampicillin resistant) primer-M13Rp1 RsiteL=Not I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' ACTCGGAGAAATTGGCGGCGAGGAATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.	
	ORGANISM	Homo sapiens	
	REFERENCE	Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
	AUTHORS	1 (bases 1 to 236) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Ie,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,N., Tan,F., Trevasaki,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.	
	TITLE	The WashU-Merck EST Project	
	JOURNAL	Unpublished (1995)	
	COMMENT	GDB: G00-410-094 Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 180 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	

<hr/>					
NCBI gi: 792196					
FEATURES					
source					
Location/Qualifiers					
1..236					
/organism="Homo sapiens"					
/clone="3f553"					
/note="human"					
BASE COUNT	38 a	75 c	68 g	47 t	8 others
ORIGIN					
<hr/>					
DB 24; Score	22;	Match 66.7%;	QryMatch 1.5%;	Pred. No. 7.53e-04;	
Matches	38;	Conservative	0;	Mismatches 19;	Indels 0; Gaps 0;
<hr/>					
Db	95	tctgtgctncttncccaccccagcagtngccctccccctccctctgttt	151		
Qy	1097	TCCTTCACTGCGTTGCCACCCAGCATGTCTCCACCCCAGCTCCACCCCTGTT	1153		
<hr/>					
RESULT	6				
LOCUS	RIC0462A	361 bp	mRNA	EST	25-MAY-1995
DEFINITION	Rice cDNA, partial sequence (C0462.1A).				

ACCESSION	D22591	EST (expressed sequence tag).
KEYWORDS	Oryza sativa (strain Nipponbare,)	Callus Callus cDNA to mRNA.
SOURCE	Oryza sativa	
ORGANISM	Oryza sativa	
REFERENCE		Eukaryotes; mitochondrial eukaryotes; Chlorophyta/Embryophyta group; Charophyta/Embryophyta group; Embryophyta; Magnoliophyta; Liliopsida; Commelinidae; Poales; Poaceae; Oryza.
AUTHORS	1 (bases 1 to 361)	
TITLE	Minobe, Y. and Sasaki, T.	
JOURNAL	Rice cDNA from callus	
COMMENT	Unpublished (1993)	
	Submitted (2-NOV-1993) to DDBJ by:	
	Yuzo Minobe	
	Dept. Rice Genome Research Program	
	National Institute of Agrobiological Resources	
	Kannondai 2-1-2	
	Tsukuba, Ibaraki	
	Japan	
	Phone: 0298-38-7441	
	Fax: 0298-38-7468	
	PROJECT = 'RGP'.	
NCBI gi:	425911	
FEATURES	Location/Qualifiers	
source	1..361	
	/organism="Oryza sativa"	
	/strain="Nipponbare"	
	/dev_stage="Callus"	
	/sequenced_mol="cDNA to mRNA"	
	/tissue_type="Callus"	
BASE COUNT	65 a 137 c 101 g 42 t 16 others	
ORGIN		
DB 32; Score	22; Match 62.8%; QryMatch 1.5%; Pred. No. 7.53e-04;	
Matches	59; Conservative 0; Mismatches 34; Indels 1; Gaps 1;	
Db	50 gatgaacaccacactctctctctccactcaactcgcctanagccgtacnccacaccca 109	
Qy	1077 GATGGACACACACCTTCCTCCTTCACCTGCTTCCGCCACCCACCGATGATCTCCACC- 1135	
Db	110 cccgcgcacactccgcctccatggcgcacacc 143	
Qy	1136 CAGCTCCACCCCTGTTCTCTGACCTTCGCC 1169	
RESULT	7	
LOCUS	R01838	405 bp mRNA EST 31-MAR-1995
DEFINITION	ye84h03.r1 Homo sapiens cDNA clone 124469 5' similar to gb:M83653	
	RED CELL ACID PHOSPHATASE 1, ISOZYME F (HUMAN);.	
ACCESSION	R01838	
KEYWORDS	EST.	
SOURCE	Human	
	human clone=124469 library=Soares fetal liver spleen INFLS	
	vector=pT73D (Pharmacia) with a modified polylinker host=DH10B	
	(ampicillin resistant) primer=M13RP1 Rsite1=Pac I Rsite2=Eco RI	
	Liver and spleen from a 20 week-post conception male fetus. 1st	
	strand cDNA was primed with a Pac I - oligo(dT) primer [5'	
	AACCTGGAGAACTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded	
	cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac	
	I and cloned into the Pac I and Eco RI sites of the modified pT73	
	vector. Library went through one round of normalization. Library	
	constructed by Bento Soares and M.Fatima Bonaldo.	
ORGANISM	Homo sapiens	
	Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;	
	Eutheria; Primates; Catarrhini; Homnidae; Homo.	

1 (bases 1 to 405)

REFERENCE
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaekis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

CONTACT: Wilson RK

WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

High quality sequence stops: 290
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 751574
Location/Qualifiers
1..405
/organism="Homo sapiens"
/clone="124469"
/note="human"

BASE COUNT 100 a 85 c 89 g 126 t 5 others

ORIGIN

DB 15; Score 22; Match 80.6%; QryMatch 1.5%; Pred. No. 7.53e-04;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Dbbt tccccagtgtaacctctctctccctaggcgctttt 213
|| ||||||| || ||||||| |||||||
Cp 1384 TCTCCAGTGTGCTGTATTCCTCTCCCAGGCCGCTTT 1349

RESULT 8

LOCUS R13668 355 bp mRNA EST 12-APR-1995

DEFINITION yf6h02.r1 Homo sapiens cDNA clone 26866 5'.

ACCESSION R13668

KEYWORDS EST.

SOURCE human clone=26866 library=Soares infant brain INTB vector=lafmid BA host=DHI10B (ampicillin resistant) primer=M13RP1 Reitel=Not I Site=Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTCGAAGAATTCCGGCCGCAGGAATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eucaryota; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 355)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaekis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT

GDB: G00-399-213
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 229
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.

NCBI gi: 766744
 Location/Qualifiers
 1..355
 /organism="Homo sapiens"
 /clone="26866"
 /note="human"

BASE COUNT 63 a 91 c 93 g 105 t 3 others
 ORIGIN

DB 18; Score 22; Match 86.7%; QryMatch 1.5%; Pred. No. 7.53e-04;
 Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 124 tctctccacagcccaactccctgttccct 153
 ||||| ||||| ||||| ||||| ||||| |||||
 Qy 1128 TCTCCACCCAGCTCCACGCCCTGTTCTT 1157

RESULT 9
 LOCUS T07692 152 bp mRNA EST 30-JUN-1993
 DEFINITION EST05582 Homo sapiens cDNA clone HFBEM78.
 ACCESSION T07692
 KEYWORDS EST.
 SOURCE Human clone=HFBEM78 library=Fetal brain, Stratagene (cat#036206)
 vector=Lambda2AP-II primer=W13-21 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; Lambda2AP-II vector, 1.0Kb
 average insert size.

ORGANISM Homo sapiens
 Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
 Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.

REFERENCE 1 (bases 1 to 152)
 AUTHORS Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
 TITLE 3400 Expressed Sequence Tags Identify Diversity of Transcripts from
 Human Brain
 JOURNAL Nature Genet. 4, 256-267 (1993)

COMMENT

Contact: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: mdadams@tigr.org.

NCBI gi: 318841
 Location/Qualifiers
 1..152
 /organism="Homo sapiens"
 /clone="HFBEM78"
 /note="Human"

BASE COUNT 52 a 21 c 51 g 24 t 4 others
 ORIGIN

SOURCE

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini, Hominoidea; Homo.
REFERENCE 1 (bases 1 to 313)

ORGANISM *Homo sapiens*

REFERENCE

AUTHORS
Hillier, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maria, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasaki, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.

TITLE
WashU-Merck EST Project

JOURNAL
Unpublished (1995)

COMMENT

Contact: Wilson RK

WashU-Merck EST Protect

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
 High quality sequence stops: 253
 Source: IMAGE Consortium, LINL
 This clone is available royalty-free through LINL ; contact the
 IMAGE Consortium (info@image.linl.gov) for further information.

NCBI si: 691597

FEATURES	Location/Qualifiers
near gr. 00100	

Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 430)

REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE
JOURNAL
COMMENT
The WashU-Merck EST Project
Unpublished (1995)

GDB: G00-400-468
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 324
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 766401
Location/Qualifiers
1..430
/organism="Homo sapiens"
/clone="28121"
/note="human"

BASE COUNT
ORIGIN
91 a 113 c 121 g 100 t 5 others

DB 18; Score 21; Match 59.1%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 212 aagaggccctaaactcccccatccagatggatggatccctccaaaggaact 266
||||| ||||| | | ||||| || | ||||| | | |||||

Cp 1406 AATAGCTCTAAATTTTACAAATCTCCAGTCTGCTGTATCCCTCCCGAGGCCT 1352

RESULT 13
LOCUS T58482 383 bp mRNA EST 09-FEB-1995
DEFINITION yb61ell.r1 Homo sapiens cDNA clone 75692 5' similar to gb:M64716
40S RIBOSOMAL PROTEIN 525 (HUMAN);.

ACCESSION T58482
KEYWORDS EST.
SOURCE human clone=75692 library=Stratagene ovary (#937217)
vector=Bluescript SK host=SOLR cells (kanamycin resistant)
primer=M13R1 Rsite=EcRI Rsite2=XhoI Cloned unidirectionally.
Primer: Oligo df. Total ovary tissue, normal 49 year old caucasian female. Average insert size: 0.8 kb; Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTGGCAGCAG-3'; 3' adaptor sequence: 5'-CTCGAGCTTTTTTTTTTTTTTT-3'.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE
JOURNAL
COMMENT
WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 197
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 660319
Location/Qualifiers
1..383
/organism="Homo sapiens"
/clone="75692"
/note="human"

BASE COUNT
ORIGIN
112 a 86 c 92 g 88 t 5 others

DB 47; Score 21; Match 75.7%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 283 cagggnccagccctttccaggagcgtcttttagtta 319
||||| ||||| | | ||||| || | |||||

Qy 123 CAGGGGAGAGGCCCCATACAGGGAGCCACTTCAGTTA 159

RESULT 14
LOCUS R07221 346 bp mRNA EST 05-APR-1995
DEFINITION yf14b06.r1 Homo sapiens cDNA clone 126803 5'.

ACCESSION R07221
KEYWORDS EST.
SOURCE human clone=126803 library=Soares fetal liver spleen INFLS
vector=pt7T3D (Pharmacia) with a modified polylinker host=DH10B
(ampicillin resistant) primer=M13R1 Rsite=Fac I Rsite2=Eco RI
Liver and spleen from a 20 week-post conception male fetus. 1st
strand cDNA was primed with a Pac I - oligo(dT) primer [5'
AACTGGAGAAATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Pac I and Eco RI sites of the modified pT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE
JOURNAL
COMMENT
The WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 211

Source: IMAGE Consortium, L1NL
This clone is available royalty-free through L1NL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 759144
Location/Qualifiers
source
1..346
/organism="Homo sapiens"
/clone="126803"
/note="human"

BASE COUNT 95 a 63 c 110 g 73 t 5 others

DB 16; Score 21; Match 88.9%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 202 tcagaagtccaggccaagctgtgtcgg 228
|||||
Cp 831 TCAGAAGTCCAGGCGCAGCAGTTCTGG 805

RESULT 15
LOCUS R12342 413 bp mRNA EST 12-APR-1995
DEFINITION yf55f05.r1 Homo sapiens cDNA clone 25883 5' similar to gb:M10058
ASTALOGLYCOPROTEIN RECEPTOR 1 (HUMAN);
ACCESSION R12342
KEYWORDS EST.
SOURCE human clone=25883 library=Soares infant brain lNIB vector=Lafmid BA
host=DH10B (ampicillin resistant) primer=M13Rpl Rsite1=Not I
Rs1te2=Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AAGTCAGACAAATTCGGCGCGCAGGAATTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 413)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

GDB: G00-398-230
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 329
Source: IMAGE Consortium, L1NL
This clone is available royalty-free through L1NL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 765418
Location/Qualifiers

source 1..413
/organism="Homo sapiens"
/clone="25883"
/note="human"
BASE COUNT 73 a 137 c 109 g 92 t 2 others
ORIGIN
DB 18; Score 21; Match 69.2%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Db 90 aggttcnctctgtcccgcgctctctgcacacccctctctctgtcagggccac 141
|||||
Cp 486 AGGGTTCCAACTGTCTCTGTGTGCGCATCATCTCCCTCCAGTAGAAGGGACAC 435
Search completed: Wed Sep 27 23:43:35 1995
Job time : 324 secs.

MAPSREL (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 27 11:41:12 1995; MasPar time 5.69 Seconds

Tabular output not generated. 423.266 Million cell updates/sec

Title: >US-08-252-491-2
Description: (1:379) from US08252491.pep
Perfect Score: 2644
Sequence: 1 MAPCK1QGRPIQGATSVRH.....STAPHPVTMYPHRNLQSET 379

Scoring table: PAM 150
Gap 11

Searched: 53402 seqs, 6354270 residues

Database: a-geneseq
1 a-gen1
2 a-gen2
3 a-gen3
4 a-gen4
5 a-gen5
6 a-gen6
7 a-gen7
8 a-gen8
9 a-gen9
10 a-gen10

Statistics: Mean 34.679; Variance 190.574; scale 0.182

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	5.0	193	3	P50300 Human erythropoietin	3.02e-01
2	131	5.0	330	4	R23076 Epo:IL-3 short, recom	3.02e-01
3	131	5.0	349	4	R23079 Epo:IL-3 Flex, recom	3.02e-01
4	131	5.0	193	2	R11892 Erythropoietin analog	3.02e-01
5	131	5.0	193	2	P70256 Sequence of human ery	3.02e-01
6	131	5.0	188	1	P81195 Erythropoietin encode	3.02e-01
7	131	5.0	193	3	P60597 Clone lambda HEPOFL13	3.02e-01
8	131	5.0	188	3	P60599 Clone lambda HEPOFL16	3.02e-01

9	131	5.0	193	2	R11858 Modified human erythr	3.02e-01
10	131	5.0	193	2	R11893 Erythropoietin analog	3.02e-01
11	131	5.0	193	2	R11891 Erythropoietin analog	3.02e-01
12	130	4.9	193	3	P50343 EPO encoded by clone	3.48e-01
13	130	4.9	193	3	P60598 Open reading frame co	3.48e-01
14	124	4.7	334	3	P83194 Sequence of a bioadhe	8.17e-01
15	118	4.5	193	2	R11859 Erythropoietin analog	1.90e+00
16	116	4.4	334	1	P82971 Bioadhesive precursor	2.50e+00
17	114	4.3	192	3	P50301 Monkey erythropoietin	3.30e+00
18	109	4.1	254	2	R10531 Prod. of DNA of pMG07	6.55e+00
19	108	4.1	372	9	R47260 Pre-pro-GDF-1.	7.51e+00
20	108	4.1	372	9	R50207 Human GDF-1(fx) morph	7.51e+00
21	108	4.1	372	9	R46750 Human morphogenic pro	7.51e+00
22	108	4.1	372	9	R60961 Human GDF-1(fx) morph	7.51e+00
23	108	4.1	372	9	R57980 Human GDF-1(fx) morph	7.51e+00
24	108	4.1	372	9	R47297 GDF-1.	7.51e+00
25	108	4.1	372	6	R34582 Human GDF-1 morphogen	7.51e+00
26	108	4.1	372	6	R33942 Morphogen GDF-1 full	7.51e+00
27	108	4.1	302	4	R23596 Recombinant hematopoi	7.51e+00
28	108	4.1	372	4	R22376 hGDF-1.	7.51e+00
29	108	4.1	121	5	R29164 PRP3.	7.51e+00
30	108	4.1	321	4	R23075 IL-3:Epo short, recom	7.51e+00
31	106	4.0	151	5	R29165 PRP3 (from genomic cl	9.84e+00
32	105	4.0	77	1	P81196 Erythropoietin encode	1.13e+01
33	104	3.9	156	2	R10532 Prod. of DNA of pMG08	1.29e+01
34	103	3.9	340	4	R23078 IL-3:Epo Flex, recom	1.47e+01
35	103	3.9	321	4	R23597 Recombinant hematopoi	1.47e+01
36	102	3.9	114	5	R29163 PRP2.	1.68e+01
37	101	3.8	533	7	R39468 hRXR-beta1.	1.92e+01
38	100	3.8	521	9	R57772 c-MEF2.	2.19e+01
39	99	3.7	322	4	R23599 Recombinant hematopoi	2.50e+01
40	99	3.7	167	3	P50298 Human recombinant ery	2.50e+01
41	99	3.7	166	2	P70398 Sequence of human ery	2.50e+01
42	99	3.7	167	3	P50299 Human recombinant ery	2.50e+01
43	99	3.7	166	4	R23593 Recombinant hematopoi	2.50e+01
44	99	3.7	303	4	R23598 Recombinant hematopoi	2.50e+01
45	97	3.7	439	5	R28150 Sugar beet chitinase	3.26e+01

ALIGNMENTS

RESULT	1
ID	P50300 standard; protein; 193 AA.
AC	P50300;
DT	01-JAN-1980 (first entry)
DE	Human erythropoietin encoded by positive clone (phage lambda-hel)
DE	Isolated from human fetal liver gene bank.
KW	Erythropoietin; red blood cell; erythrocyte; anaemia; blood;
KW	disorder; ss; phage lambda-hel; gene bank.
OS	Homo sapiens.
PN	W08502610-A.
PD	20-JUN-1985.
PF	11-DEC-1984; U02021.
PR	13-DEC-1983; US-561024.
PR	21-FEB-1984; US-582185.
PR	28-SEP-1984; US-655841.
PR	30-NOV-1984; US-675298.
PA	(KIRI-) KIRIN-AMGEN INC.
DR	N-PSDB; N50347.
PT	New polypeptide having properties of erythropoietin - is prepd.
PT	by cultivation of transformed eucaryotic or procaryotic host
PS	Disclosure; Page 43; 113pp; English.
CC	Human erythropoietin encoded by a sequence encoded by this phage

ID R11892 standard; Protein; 193 AA.
AC R11892;
DT 22-JUL-1991 (first entry)
DE Erythropoietin analogue [Thr1125].
KW Human erythropoietin; EPO; isoform; analogue; haematocrit;
KW sialic acids; glycosylation.
OS Homo sapiens.
PN W09105867-A.
PN EP-428267-A.
PD 02-MAY-1991.
PF 09-OCT-1990; U05758.
PR 13-OCT-1989; US-421444.
PA (AMGE-) AMGEN INC.
PI Strickland TW, Byrne TE, Elliott SG;
DR WPI; 91-148745/20.
DR WPI; 91-150265/21.
PT Recombinant erythropoietin iso-forms and purificn. - increase
PT haematocrit levels in mammals and contg. specific number of
PT sialic acids
PS Claim 33; Page 45; 60pp; English
CC The analogue was constructed to add an O-glycosylation site
CC at Thr125 (amino acid 154 in this sequences) and was produced by
CC expression of DNA obtained by site-directed mutagenesis of DNA
CC encoding EPO. The analogue has a higher sialic acid content than
CC human EPO and has increased biological activity.
CC It can be used to treat mammals to cause bone marrow cells to
CC increase prodn. of reticulocytes and red blood
CC thereby increasing haematocrit levels.
CC See also R11859, R11891-93.
SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;
Db 12 lllslslslplglpvgaprrlclcdsrlylleakeaenittgcaehcslnenitvptd 71
|||: : l::: ||:| ||:|:|: |:: : : l::: ||: : : |
Qy 29 LLLAAMLLAVARLTUSPVAPACDRLNKLRLDRLHSLHSLRSLQCQPDVPLSPVLLPAV 88
:: || |:: : ||:| ||:|:|: |:: : : l::: ||: : : |
Db 72 knfyawkrmevgqgavevvgqlllseavlrqall 108
:: || |:: : ||:| ||:|:|: |:: : : l::: ||: : : |
Qy 89 DFLSGEMKRTQTEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 5
ID P70256 standard; protein; 193 AA.
AC P70256;
DT 19-FEB-1991 (first entry)
DE Sequence of human erythropoietin (EPO).
KW Renal anaemia therapy; hormone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..27
FT /label=SIGNAL 28..193
FT Protein 81..97
FT Region
FT /note="Fragment that probe N70361 is based on"
PN EP-232034-A.
PD 12-AUG-1987.
PF 19-JAN-1987; 300399.
PR 23-JAN-1986; JP-012868.
PA (SUMO) SUMITOMO CHEM IND KK.
PA (SUMI-) SUMITOMO SEIYAKU KK.
PI Yanagi H, Ogawa I, Okamoto M, Hozumi T, Soga A, Yoshima T,
Tsuchumi M.

DR WPI; 87-223006/32.
DR N-PSDB; N70360, N70361.
PT Human erythropoietin prodn. - by culturing human cells, esp.
PT Namalwa cells, transfected with DNA encoding human erythropoietin
PS Disclosure; Fig 1; 22pp; English.
CC A cDNA library was prepd. from the poly (A) RNA, which was isolated
CC from the erythropoietin-producing human hepatoma cell Hp-1. The cDNA
CC library was screened using the probes given in N70361 and N70362. A
CC plasmid (named as p58-A20) was isolated. The nucleotide sequence of
CC the cDNA obtained from this clone is shown in N70360.
SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;
Db 12 lllslslslplglpvgaprrlclcdsrlylleakeaenittgcaehcslnenitvptd 71
|||: : l::: ||:| ||:|:|: |:: : : l::: ||: : : |
Qy 29 LLLAAMLLAVARLTUSPVAPACDRLNKLRLDRLHSLHSLRSLQCQPDVPLSPVLLPAV 88
:: || |:: : ||:| ||:|:|: |:: : : l::: ||: : : |
Db 72 knfyawkrmevgqgavevvgqlllseavlrqall 108
:: || |:: : ||:| ||:|:|: |:: : : l::: ||: : : |
Qy 89 DFLSGEMKRTQTEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 6
ID P81195 standard; protein; 188 AA.
AC P81195;
DT 20-NOV-1990 (first entry)
DE Erythropoietin encoded by EPO 140B.
KW EPO; erythropoietin; anaemia; renal failure.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label=leader sequence 23..188
FT Protein
FT /label=EPO
FN EP-267678-A.
PD 18-MAY-1988.
PF 15-SEP-1987; 308130.
PR 15-SEP-1986; US-907369.
PA (INTE-) Integrated genetics.
PI Beck AK, Withy RM, Zabrecky JR, Massiello NC;
DR WPI; 88-134531/20.
DR N-PSDB; N81554.
PT Recombinant human erythropoietin - produced by a transformed rodent
PT capable of producing N-linked and O-linked glycosylated human
PT erythropoietin.
PS Disclosure; p; English.
CC EPO 104B was one of four positive clones isolated from a cDNA
CC library prepd. from mRNA extracted from a human foetus of about 20
CC wk. gestation. The clone was identified using two probes, EPO1 and
CC EPO2 based on the published sequence of EPO (Nature (1985) Vol.313,
CC p.806). The sequence between nucleotides 63 and 724 has 100% homo-
CC logy with the published sequence. It encodes the 166 AAs of the
CC mature EPO protein and 22 AAs of the leader sequence. This clone
CC and a second, EPO 125, were used to construct a full length clone
CC which was expressed in rodent epithelial cells.
CC See also P81196.
SQ Sequence 188 AA;

DB 1; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;
Db 7 lllslslslplglpvgaprrlclcdsrlylleakeaenittgcaehcslnenitvptd 66

Qy 29 LLLAAMLLAVARLTLSVPACDPRLINKLRSHLLHSRLSQCPDVPDPLSLPVLPAV 88
 |||: : ||:: |::| |::| |::| |::| : : ||: : |: : |:

Db 67 kmfyawkrmevggqavewgglallseavlrqgall 103
 :: || |::| |::| |::| |::| |::| |::|

Qy 89 DFSLGEMKQTQEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 7

ID P60597 standard; protein; 193 AA.
 AC P60597;
 DE 01-JAN-1980 (first entry)
 DE Clone lambda HEPOFL13 encoding human erythropoietin.
 KW Erythropoietin; lambda HEPOFL13; recombinant plasmid vector; anaemia;
 KW mammal cell culture; 3T3; CHO; Chinese hamster ovary; ss.
 OS Homo sapiens.
 PN W08603520-A.
 PD 19-JUN-1986.
 PF 03-DEC-1985; U02405.
 PR 04-DEC-1984; US-677813.
 PR 03-JAN-1985; US-688622.
 PR 22-JAN-1985; US-693258.
 PA (GENE-) GENETICS INST INC.
 PA (FRIT/) FRITSCHE E.
 PI Fritsch E, Hewick RM, Jacobs K;
 DR WPI; 86-169459/26.
 DR N-PSDB; N60513.
 PT Prodn. of human cDNA clone expressing erythropoietin - for mass
 prodn. of erythropoietin, useful for treating anaemia
 PS Disclosure; Page 7; 61pp; English.
 CC A recombinant plasmid vector expressing this clone is expressed in e.
 CC g 3T3 or CHO cell cultures. The produced erythropoietin is useful
 CC for treatment of anaemia, especially renal anaemia. The cloned gene
 CC expresses high levels of the protein and thus provides a means of
 CC mass production. See also N60514-21 and P60598-99.
 CC mass production. See also N60514-21 and P60598-99.
 SQ Sequence 193 AA;

DB 3; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 lllslslplglpvgaprrlclcdsrlylleakeaenittgcaehcslnenitvpdt 71
 |||: : ||:: |::| |::| |::| |::| : : ||: : |: : |:

Qy 29 LLLAAMLLAVARLTLSVPACDPRLINKLRSHLLHSRLSQCPDVPDPLSLPVLPAV 88

Db 72 kmfyawkrmevggqavewgglallseavlrqgall 108

Qy 89 DFSLGEMKQTQEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 8

ID P60599 standard; protein; 188 AA.
 AC P60599;
 DE 01-JAN-1980 (first entry)
 DE Clone lambda HEPOFL16 encoding human erythropoietin.
 KW Erythropoietin; lambda HEPOFL16; recombinant plasmid vector; anaemia;
 KW mammal cell culture; 3T3; CHO; Chinese hamster ovary; ss.
 OS Homo sapiens.
 PN W08603520-A.
 PD 19-JUN-1986.
 PF 03-DEC-1985; U02405.
 PR 04-DEC-1984; US-677813.
 PR 03-JAN-1985; US-688622.
 PR 22-JAN-1985; US-693258.
 PA (GENE-) GENETICS INST INC.

PA (FRIT/) FRITSCHE E.
 PI Fritsch E, Hewick RM, Jacobs K;
 DR WPI; 86-169459/26.
 DR N-PSDB; N60519.

PT Prodn. of human cDNA clone expressing erythropoietin - for mass
 prodn. of erythropoietin, useful for treating anaemia
 PS Disclosure; Page 20; 61pp; English.
 CC A recombinant plasmid vector expressing this clone is expressed in e.
 CC g 3T3 or CHO cell cultures. The produced erythropoietin is useful
 CC for treatment of anaemia, especially renal anaemia. The cloned gene
 CC expresses high levels of the protein and thus provides a means of
 CC mass production. See also N60513-18, N60520-21 and P60598.
 CC mass production. See also N60513-18, N60520-21 and P60598.
 SQ Sequence 188 AA;

DB 3; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 7 lllslslplglpvgaprrlclcdsrlylleakeaenittgcaehcslnenitvpdt 66
 |||: : ||:: |::| |::| |::| |::| : : ||: : |: : |:

Qy 29 LLLAAMLLAVARLTLSVPACDPRLINKLRSHLLHSRLSQCPDVPDPLSLPVLPAV 88

Db 67 kmfyawkrmevggqavewgglallseavlrqgall 103

Qy 89 DFSLGEMKQTQEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 9

ID R11838 standard; protein; 193 AA.
 AC R11838;
 DT 24-JUL-1991 (first entry)
 DE Modified human erythropoietin.
 KW EPO; anaemia.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Protein 28..193
 FT /label= Mature modified EPO
 FT Misc.difference 51
 FT /label= May be absent or any AA except Asn
 FT Misc.difference 65
 FT /label= May be absent or any AA except Asn
 FT Misc.difference 110
 FT /label= May be absent or any AA except Asn
 FT EP-427189-A.
 PD 15-MAY-1991.
 PF 06-NOV-1990; 121187.
 PR 07-NOV-1989; JP-289143.
 PA (SNOW) SNOW BRAND MILK PRODUCTS.
 PI Yamaguchi K, Ueda M;
 DR WPI; 91-141727/20.
 PT New human erythropoietin(s) modified at positions 24, 38 and 83 -
 PT having improved biological activity for stimulating erythrocyte
 PT differentiation and treating anaemia.
 PS Disclosure; Page 8-9; 14pp; English.
 CC Modified EPO has a reduced number of sialic acid binding sites, and
 CC the galactose residue is not exposed. By stripping away these sugars
 CC from the glycoprotein, the affinity of the EPO to its receptor is
 CC greatly enhanced. Thus, the modified EPO shows a higher activity and
 CC is useful in the treatment of anaemia.
 SQ Sequence 193 AA;

DB 2; Score 131; Match 23.0%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 23; Conservative 32; Mismatches 45; Indels 0; Gaps 0;

Db 12 lllslslplglpvgaprrlclcdsrlylleakeaenittgcaehcslnenitvpdt 71

Qy 29 LLLAAMLLAVARLTSSVPACDPRLINKLRDLSHLLSRSLSCQPDVPLSIPVLLPAV 88
 Db 72 kmfyawkrmevgqavewgqlallseavlrqgallvqs 111
 Qy 89 DFSLGEKMTQTEQSKAQDILGAVSLLEGVMAARGQLS 128

RESULT 10

ID R11893 standard; Protein; 193 AA.
 AC R11893;
 DT 22-JUL-1991 (first entry)
 DE Erythropoietin analogue [Pro]124, [Thr]125.
 KW Human erythropoietin; EPO; isoform; analogue; haematocrit;
 KW sialic acids; glycosylation.
 OS Homo sapiens.
 PN W09105867-A.
 PN EP-428267-A.
 PD 02-MAY-1991.
 PF 09-OCT-1990; U05758.
 PR 13-OCT-1989; US-421444.
 PA (AMGE-) AMGEN INC.
 PI Strickland TW, Byrne TE, Elliott SG;
 DR WPI; 91-148745/20.
 DR WPI; 91-150265/21.
 PT Recombinant erythropoietin iso-forms and purificn. - increase
 PT haematocrit levels in mammals and contg. specific number of
 PT sialic acids
 PS Claim 33; Page 45; 60pp; English
 CC The analogue was constructed to add an O-glycosylation site
 CC at Thr125 (amino acid 154 in this sequences) and was produced by
 CC expression of DNA obtained by site-directed mutagenesis of DNA
 CC encoding EPO. The analogue has a higher sialic acid content than
 CC human EPO and has increased biological activity.
 CC It can be used to treat mammals to cause bone marrow cells to
 CC increase prodn. of reticulocytes and red blood
 CC thereby increasing haematocrit levels.
 CC See also R11859, R11891-93.
 SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 LLLSLLSLPLGLPVGAPRLICDSRVLYLLEAEENITGCAEHCALNENITVPTD 71
 Qy 29 LLLAAMLLAVARLTSSVPACDPRLINKLRDLSHLLSRSLSCQPDVPLSIPVLLPAV 88
 Db 72 kmfyawkrmevgqavewgqlallseavlrqgall 108
 Qy 89 DFSLGEKMTQTEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 11

ID R11891 standard; Protein; 193 AA.
 AC R11891;
 DT 22-JUL-1991 (first entry)
 DE Erythropoietin analogue [Asn]125, [Ser]127.
 KW Human erythropoietin; EPO; isoform; analogue; haematocrit;
 KW sialic acids; glycosylation.
 OS Homo sapiens.
 PN W09105867-A.
 PN EP-428267-A.
 PD 02-MAY-1991.
 PF 09-OCT-1990; U05758.

PR 13-OCT-1989; US-421444.
 PA (AMGE-) AMGEN INC.
 PI Strickland TW, Byrne TE, Elliott SG;
 DR WPI; 91-148745/20.
 DR WPI; 91-150265/21.
 PT Recombinant erythropoietin iso-forms and purificn. - increase
 PT haematocrit levels in mammals and contg. specific number of
 PT sialic acids
 PS Claim 33; Page 45; 60pp; English
 CC The analogue was constructed to add an N-glycosylation site
 CC at Asn125 (amino acid 154 in this sequences) and was produced by
 CC expression of DNA obtained by site-directed mutagenesis of DNA
 CC encoding EPO. The analogue has a higher sialic acid content than
 CC human EPO and has increased biological activity.
 CC It can be used to treat mammals to cause bone marrow cells to
 CC increase prodn. of reticulocytes and red blood
 CC thereby increasing haematocrit levels.
 CC See also R11859, R11891-93.
 SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 LLLSLLSLPLGLPVGAPRLICDSRVLYLLEAEENITGCAEHCALNENITVPTD 71
 Qy 29 LLLAAMLLAVARLTSSVPACDPRLINKLRDLSHLLSRSLSCQPDVPLSIPVLLPAV 88
 Db 72 kmfyawkrmevgqavewgqlallseavlrqgall 108
 Qy 89 DFSLGEKMTQTEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 12

ID P50343 standard; Protein; 193 AA.
 AC P50343;
 DT 10-MAR-1992 (first entry)
 DE EPO encoded by clone lambda-HEPOFL13.
 KW Erythropoietin.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..27
 FT /label= signal_sequence
 FT Protein 27..193
 FT /label= mature_EPO
 PN DK8406107-A.
 PD 16-AUG-1985.
 PF 27-DEC-1984; 281862.
 PR 27-DEC-1983; US-565627.
 PR 27-DEC-1983; US-566057.
 PR 04-DEC-1984; US-677813.
 PR 21-DEC-1984; NO-005186.
 PR 21-DEC-1984; ZA-010034.
 PR 22-JAN-1985; US-693258.
 PR 22-JAN-1985; US-688622.
 PA (GENE-) GENETICS INST.
 PI Kaufman RJ.
 DR WPI; 85-318061/51.
 DR N-PSDB; N50443.
 PT Vector system for introducing heterologous DNA into eukaryotic
 PT cells - comprises prod. gene and accessory DNA for enhanced
 PT expression of heterologous protein by the cells.
 PS Disclosure; Fig 13; 62pp; Danish.
 CC See also US4740461 88.04.26 (8819) (first major country equivalent).
 CC The sequence is encoded by clone lambda-HEPOFL13 which contains

Qy 89 DFLGKTKTQTEQSKAQDILGAVSLLEGVMAARGQL 125

PPT Recombinant erythropoietin iso-forms and purificn. - increase
PPT haematocrit levels in mammals and contg. specific number of

PT sialic acids
 PS Claim 33; Page 45; 60pp; English
 CC The [Asn]69 analogue was constructed to add an N-glycosylation site
 CC at Asn69 (amino acid 96 in this sequences) and was produced by
 CC expression of DNA obtained by site-directed mutagenesis of DNA
 CC encoding EPO. The analogue has a higher sialic acid content than
 CC human EPO and has increased biological activity.
 CC It can be used to treat mammals to cause bone marrow cells to
 CC increase prodn. of reticulocytes and red blood
 CC thereby increasing haematocrit levels.
 CC See also R11859, R11891-93.
 SQ Sequence 193 AA;

DB 2; Score 118; Match 21.6%; QryMatch 4.5%; Pred. No. 1.90e+00;
 Matches 21; Conservative 33; Mismatches 43; Indels 0; Gaps 0;

Db 12 lllslslplglplgprlricdsrvlrylleakeanittgcaehcslnenitvpdt 71
 |||: : l::: ||: ||: ||: ||: : : ||: : ||: : ||

Qy 29 LLLAAMLLAVARLTSSVPAPACDPRLANKLRDShLLHSLRSQCpDVPdLSIPVLLPAV 88
 |||: : l::: ||: ||: ||: ||: : : ||: : ||: : ||

Db 72 knfyawkrmvvgqavevwqglanlseavlrqgall 108
 :: || ||: ||: ||: ||: ||: ||: ||

Qy 89 DFLGKWKQTQEQSKAQDILGAVSLLEGVMAARGQL 125
 :: || ||: ||: ||: ||: ||: ||: ||

Search completed: Wed Sep 27 11:41:31 1995
 Job time : 19 secs.


```

SUMMARY          #length 356 #molecular-weight 37835 #checksum 9983
DB 12; Score      2482; Match 100.0%; QryMatch 93.9%; Pred. No. 0.00e+00;
Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 meltlllvaamlavartlaspvaccdrlnklldshlherlscqdvplspv 60
Qy 24 MELTDLLAAMLLAVARUTLSSVPACDPRLNKLGRDHLHLSRLSQCPDVPDLSPV 83

Db 61 llpavdflgskwtqteqskadilgavelllegvmaargqlpcslslqqlsgqvrl 120
Qy 84 LLPAVDFSLEGEKWTQTEQSKAQDILGAVSLLEGVMAARGQLEPSCLSLLGQLSQVRL 143

Db 121 llgalqllgtqlpqrtrtkhkdnaiflqlgrkvrlflllvegptlcvrtppt 180
Qy 144 LLGALQLLGTQLPQGRTRTKHKNALFLSLIQQLRGKVRFLLLVEGPTLCVRRTPPT 203

Db 181 avpsstsqlltlnkfmrtsqletnfsvartagqllslrlqgrfvkitpqdlnqtsrs 240
Qy 204 AVPSSTSQLTLNKFMRTSGLTWNFSVARTAGPGLLSRLQGRFVKITPQCLNQTSRS 263

Db 241 pvqlsgylnrthgvpngthglfagtslqtleadspgafnkgslafnlqgqlppspela 300
Qy 264 PVQLSGYLNRTHGVPNGTHGLFAGTSLQTLLEASDSPGAFNKGSLAFNLQGGPPSPSIA 323

Db 301 pdgthtppspalpntthgspqqlhplfdpdpstmpnstaphvmtymphprnlsoet 356
Qy 324 PDGHTPPSPALPNTTHGSPQQLHPLFDPDSTTMPNSTAPHVMTYMPHPRNLSOET 379

RESULT 2
ENTRY c-MPL ligand - human
TITLE c-MPL ligand - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Dec-1994; #sequence_revision 10-Dec-1994; #text_change 10-Dec-1994

ACCESSIONS S45331
REFERENCE S45331
#authors de Sauvage, F.J.; Hass, P.E.; Spencer, S.D.; Malloy, B.E.; Gurney, A.L.; Spencer, S.A.; Darbonne, W.C.; Henzel, W.J.; Wong, S.C.; Kuang, W.J.; Oles, K.J.; Hultgren, B.; Solberg Jr., L.A.; Goeddel, D.V.; Eaton, D.L.
#journal Nature (1994) 369:533-538
#title Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand.
#accession S45331
#status preliminary
#residues 1-353 ##label DEA
SUMMARY #length 353 #molecular-weight 37822 #checksum 1597

DB 12; Score      1748; Match 71.6%; QryMatch 66.1%; Pred. No. 2.71e-221;
Matches 255; Conservative 47; Mismatches 49; Indels 5; Gaps 4;

Db 1 meltlllvaamlavartlaspvaccdrlnklldshlherlscqdvplspv 60
Qy 24 MELTDLLAAMLLAVARUTLSSVPACDPRLNKLGRDHLHLSRLSQCPDVPDLSPV 83

Db 61 llpavdflgskwtqteqskadilgavelllegvmaargqlpcslslqqlsgqvrl 120
Qy 84 LLPAVDFSLEGEKWTQTEQSKAQDILGAVSLLEGVMAARGQLEPSCLSLLGQLSQVRL 143

Db 121 llgalqllgtqlpqrtrtkhkdnaiflqlgrkvrlflllvegptlcvrtppt 180
Qy 144 LLGALQLLGTQLPQGRTRTKHKNALFLSLIQQLRGKVRFLLLVEGPTLCVRRTPPT 203

```

```

Db 181 avpsstsqlltlnkfmrtsqletnfsvartagqllslrlqgrfvkitpqdlnqtsrs 239
Qy 204 AVPSSTSQLTLNKFMRTSGLTWNFSVARTAGPGLLSRLQGRFVKITPQCLNQTSRS 263

Db 240 ldqibgylnrthgvpngthglfagtslqtleadspgafnkgslafnlqgqlppspela 299
Qy 264 PVQISGYLNRTHGVPNGTHGLFAGTSLQTLLEASDSPGAFNKGSLAFNLQGGPPSPSIA 323

Db 300 pteqytlfpptlptp--vv-qlhplldpdeaptptptapllntsythcmleq 352
Qy 324 PDGH-TTEPFPSPALPTEGSPQLHPLFDPDSTTMPNSTAPHVMTYMPHPRNLSQE 378

RESULT 3
ENTRY A55530 #type complete
TITLE megakaryocyte growth and development factor, long form - human
ALTERNATE_NAMES MPL ligand, long form
ORGANISM #formal_name Homo sapiens #common_name man
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 31-Mar-1995

ACCESSIONS A55530
REFERENCE A55530
#authors Chang, M.; McNinch, J.; Basu, R.; Shutter, J.; Heu, R.; Perkins, C.; Mar, V.; Suggs, S.; Welcher, A.; Li, L.; Lu, H.; Bartley, T.; Hunt, P.; Martin, F.; Samal, B.; Bogenberger, J.
#journal J. Biol. Chem. (1995) 270:511-514
#title Cloning and characterization of the human megakaryocyte growth and development factor (MGDF) gene.
#accession A55530
#status preliminary
#molecule_type DNA
#residues 1-286 ##label CHA
#cross-references GB:U17071
#note sequence not compared to nucleotide translation
GENETICS
#gene MGDF
#map_position 3q26.3
KEYWORDS alternative splicing; cytokine
SUMMARY #length 286 #molecular-weight 31544 #checksum 6126

DB 9; Score      983; Match 83.6%; QryMatch 37.2%; Pred. No. 1.27e-113;
Matches 133; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

Db 1 meltlllvaamlavartlaspvaccdrlnklldshlherlscqdvplspv 60
Qy 24 MELTDLLAAMLLAVARUTLSSVPACDPRLNKLGRDHLHLSRLSQCPDVPDLSPV 83

Db 61 llpavdflgskwtqteqskadilgavelllegvmaargqlpcslslqqlsgqvrl 120
Qy 84 LLPAVDFSLEGEKWTQTEQSKAQDILGAVSLLEGVMAARGQLEPSCLSLLGQLSQVRL 143

Db 121 llgalqllgtqlpqrtrtkhkdnaiflqlgrkvrlflllvegptlcvrtppt 159
Qy 144 LLGALQLLGTQLPQGRTRTKHKNALFLSLIQQLRGKVRFLLLVEGPTLCVRRTPPT 182

RESULT 4
ENTRY S28148 #type complete
TITLE erythropoietin - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change

```



```

FEATURE
1-27
28-193
34-188,56-60
51,65,110
153
SUMMARY
#domain signal sequence #status Predicted #label SIG\
#product erythropoietin #status experimental #label MAT\
#disulfide bonds #status experimental\
#binding site carbohydrate (Asn) (covalent) #status
experimental\
#binding site carbohydrate (Ser) (covalent) #status
experimental
#length 193 #molecular-weight 21307 #checksum 6371

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 2.41e-02;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 LLLSLSLPLGLVGPAPPLICDRVLYLLEAKEENITGCAESLNNENITVPTD 71
||||:|::| |::|::|::| |::|::|::| |::|::|::| |::|::|::|
Qy 29 LLLAAMLLAVARUTLSSVPACDPRLNKLRLDRSHLLHSRLSQCPDVPDPLSIPVLLPAV 88

Db 72 kvnfyawkrmevgqqavevwqglallseavlrgqall 108
::|| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 89 DFSLGEKKTQTEQSKAQDILGAVSLLLEGVMAARGQL 125

RESULT 7
ENTRY S24407 #type complete
TITLE formin isoform IV - mouse
ORGANISM #formal name Mus musculus #common name house mouse
DATE 19-Feb-1994; #sequence_revision 19-Feb-1994; #text_change
19-Feb-1994
ACCESSIONS S24407
REFERENCE S24407
#authors Jackson-Grusby, L.; Kuo, A.; Leder, P.
#journal Genes Dev. (1992) 6:29-37
#title A variant limb deformity transcript expressed in the
embryonic mouse limb defines a novel formin.
#cross-references MUID:92112033
#accession S24407
#status preliminary
#residues 1-1206 ##label JAC
##cross-references EMBL:X62379
SUMMARY #length 1206 #molecular-weight 133463 #checksum 501

DB 12; Score 117; Match 40.0%; QryMatch 4.4%; Pred. No. 5.28e-01;
Matches 20; Conservative 12; Mismatches 16; Indels 2; Gaps 2;

Db 658 ipppplpp-glglppappipvcvppppppp-pppvtvppsdgppp 705
:|::|::|::| |::|::|::|::|::|::|::|::|::|::|::|::|
Qy 316 LPSPSLAPDGTTPFPSPALPTTHGSPQQLHPLFPDPTTMTNPSTAPHP 365

RESULT 8
ENTRY S11515 #type complete
TITLE formin - mouse
ORGANISM #formal name Mus musculus #common name house mouse
DATE 22-Jan-1994; #sequence_revision 22-Jan-1994; #text_change
22-Jan-1994
ACCESSIONS S11515
REFERENCE S11515
#authors Woychik, R.P.; Maas, R.L.; Zeller, R.; Vogt, T.F.; Leder, P.
#journal Nature (1990) 346:850-853
#title 'Formins': proteins deduced from the alternative transcripts
of the limb deformity gene.
#accession S11515
#status preliminary
#residues 1-1468 ##label WQY

```

```

##cross-references EMBL:X53599
SUMMARY #length 1468 #molecular-weight 163808 #checksum 6099

DB 12; Score 117; Match 40.0%; QryMatch 4.4%; Pred. No. 5.28e-01;
Matches 20; Conservative 12; Mismatches 16; Indels 2; Gaps 2;

Db 884 ipppplpp-glglppappipvcvppppppp-pppvtvppsdgppp 931
:|::|::|::| |::|::|::|::|::|::|::|::|::|::|::|::|
Qy 316 LPSPSLAPDGTTPFPSPALPTTHGSPQQLHPLFPDPTTMTNPSTAPHP 365

RESULT 9
ENTRY JQ0173 #type complete
TITLE erythropoietin precursor - crab-eating macaque
ORGANISM #formal name Macaca fascicularis #common name crab-eating
macaque
DATE 07-Sep-1990; #sequence_revision 07-Sep-1990 #text_change
03-Mar-1995
ACCESSIONS JQ0173
REFERENCE JQ0173
#authors Lin, F.K.; Lin, C.H.; Lai, P.H.; Browne, J.K.; Egrie, J.C.;
Smalling, R.; Fox, G.M.; Chen, K.K.; Castro, M.; Suggs, S.
#journal Gene (1986) 44:201-209
#title Monkey erythropoietin gene: cloning, expression and
comparison with the human erythropoietin gene.
#cross-references MUID:87055236
#accession JQ0173
#molecule_type mRNA
##residues 1-192 ##label LIN
##experimental_source kidney
COMMENT This protein is the principal hormone involved in the regulation of
erythrocyte differentiation and the maintenance of a
physiological level of circulating erythrocyte mass.
CLASSIFICATION #superfamily erythropoietin
KEYWORDS erythropoiesis; glycoprotein; hormone
FEATURE
1-24
25-192
51,65,110
SUMMARY #length 192 #molecular-weight 21113 #checksum 5284

DB 4; Score 114; Match 22.0%; QryMatch 4.3%; Pred. No. 9.99e-01;
Matches 22; Conservative 32; Mismatches 46; Indels 0; Gaps 0;

Db 12 LLLSLSLPLGLVGPAPPLICDRVLYLLEAKEENITGCAESLNNENITVPTD 71
||||:|::| |::|::|::| |::|::|::| |::|::|::| |::|::|::|
Qy 29 LLLAAMLLAVARUTLSSVPACDPRLNKLRLDRSHLLHSRLSQCPDVPDPLSIPVLLPAV 88

Db 72 kvnfyawkrmevgqqavevwqglallseavlrgqavllans 111
::|| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 89 DFSLGEKKTQTEQSKAQDILGAVSLLLEGVMAARGQLPS 128

RESULT 10
ENTRY A24901 #type complete
TITLE erythropoietin precursor - mouse
ORGANISM #formal name Mus musculus #common name house mouse
DATE 19-May-1989; #sequence_revision 19-May-1989 #text_change
18-Jun-1993
ACCESSIONS A24901
REFERENCE A24901
#authors McDonald, J.D.; Lin, F.K.; Goldwasser, E.
#journal Mol. Cell. Biol. (1986) 6:842-848

```



```

ORGANISM      #formal name Homo sapiens #common name man
DATE          30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
              23-Mar-1993
ACCESSIONS    S01387
REFERENCE     S01387
#authors      Sillekens, P.T.G.; Beijer, R.P.; Habets, W.J.; van Venrooij,
              W.J.
#journal      Nucleic Acids Res. (1988) 16:8307-8321
#title        Human U1 snRNP-specific C protein: complete cDNA and protein
              sequence and identification of a multigene family in
              mammals.
#cross-references MUID:88335591
#accession     S01387
#molecule_type mRNA
#residues      1-159 #label SIL
#cross-references EMBL:X12517
SUMMARY       #length 159 #molecular-weight 17394 #checksum 3819

DB 10; Score 113; Match 31.4%; QryMatch 4.3%; Pred. No. 1.23e+00;
Matches 22; Conservative 18; Mismatches 27; Indels 3; Gaps 3;

Db 51 dktaafqgkkipptfseappagamipppslpqpprrpgmmaphmgppmmmmgppp 110
|::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 297 DISPAFNKGLAFNLQGLFSPSLAPDGH-PPPPSPAL-PTTH-GSPPLHLFPDP 353

Db 111 pgmmvpgpap 120
: ||::|
Qy 354 STTMNSTAP 363

```

```

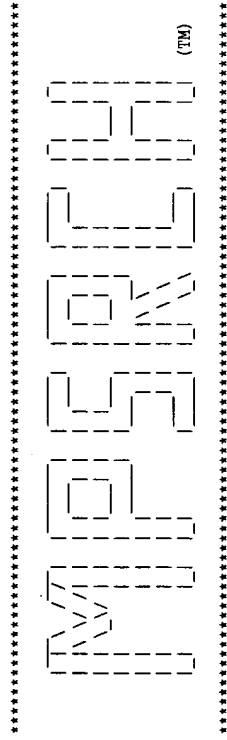
RESULT 15
ENTRY   S06733 #type complete
TITLE   hydroxyproline-rich glycoprotein precursor - common tobacco
ORGANISM #formal name Nicotiana tabacum #common name common tobacco
DATE     28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
        31-Dec-1993
ACCESSIONS S06733
REFERENCE   S06733
#authors    Keller, B.; Lamb, C.J.
#journal     Genes Dev. (1989) 3:1639-1646
#title       Specific expression of a novel cell wall hydroxyproline-rich
              glycoprotein gene in lateral root initiation.
#cross-references MUID:90128263
#accession     S06733
#molecule_type DNA
#residues      1-620 #label KEL
#cross-references EMBL:X13885
KEYWORDS    glycoprotein
SUMMARY     #length 620 #molecular-weight 65406 #checksum 1955

DB 7; Score 112; Match 33.3%; QryMatch 4.2%; Pred. No. 1.52e+00;
Matches 21; Conservative 16; Mismatches 23; Indels 3; Gaps 3;

Db 369 lpppppspp-pppsfpppppyeqpppppayaplpapptyspppptyspppptyagpp 427
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 316 LPPPSLAPDGHPT-PPSPALPTTHGSPPLHLFPDPST-TMPNSTAPHPVTWYPHPR 373

Db 428 plp 430
|:
Qy 374 NLS 376

```



Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 27 11:39:31 1995; MacPar time 9.21 Seconds
631.298 Million cell updates/sec
Tabular output not generated.

Title: >US-08-252-491-2
Description: (1:379) from US08252491.pep
Perfect Score: 2644
Sequence: 1 MAPCKIQGRGICQATSVRH.....STAPHPVTMYPHRLNSQET 379

Scoring table: PAM 150
Gap 11

Searched: 43470 seqs, 15335248 residues

Database: swiss-prot31
1 SPT1
2 SPT2
3 SPT3
4 SPT4
5 SPT5
6 SPT6
7 SPT7
8 SPT8

Statistics: Mean 49.096; Variance 112.462; scale 0.437

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	2482	93.9	356	7	TPO_MOUSE THROMBOPOIETIN PRECUR	0.00e+00
2	1748	66.1	353	7	TPO_HUMAN THROMBOPOIETIN PRECUR	1.52e-287
3	148	5.6	192	3	EPO_MOUSE ERYTHROPOIETIN PRECUR	5.93e-06
4	145	5.5	192	3	EPO_RAT ERYTHROPOIETIN PRECUR	1.46e-05
5	136	5.1	188	3	EPO_FELCA ERYTHROPOIETIN PRECUR	2.06e-04
6	134	5.1	175	3	EPO_CANFA ERYTHROPOIETIN PRECUR	3.68e-04
7	131	5.0	193	3	EPO_HUMAN ERYTHROPOIETIN PRECUR	8.68e-04
8	128	4.8	194	3	EPO_SHEEP ERYTHROPOIETIN PRECUR	2.03e-03
9	117	4.4	1468	3	FORM_MOUSE FORMIN (LIMB DEFORMIT	4.14e-02
10	117	4.4	1206	3	FORM_MOUSE FORMIN 4 (LIMB DEFORM	4.14e-02

11	114	4.3	192	3	EPO_MACFA	ERYTHROPOIETIN PRECUR	9.17e-02
12	113	4.3	159	6	RUIC_HUMAN	U1 SMALL NUCLEAR RIBO	1.19e-01
13	112	4.2	620	3	EXTN_TOBAC	EXTENSIN PRECURSOR (C	1.55e-01
14	108	4.1	372	3	GDF1_HUMAN	EMBRYONIC GROWTH FACT	4.32e-01
15	107	4.0	282	3	GDA3_WHEAT	ALPHA/BETA-GLIADIN PR	5.57e-01
16	106	4.0	744	1	CA18_HUMAN	COLLAGEN ALPHA 1(VIII	7.15e-01
17	106	4.0	1874	6	POLR_KYMWJ	RNA REPLICASE POLYPRO	7.15e-01
18	105	4.0	1213	3	FORM_CHICK	FORMIN (LIMB DEFORMIT	9.18e-01
19	105	4.0	309	4	HXA4_CHICK	HOMEOBOX PROTEIN HOX-	9.18e-01
20	104	3.9	400	8	YL37_CAEEL	HYPOTHETICAL 45.5 KD	1.18e+00
21	104	3.9	485	7	SSGP_VOLCA	SULFATED SURFACE GLYC	1.18e+00
22	104	3.9	159	6	RUIC_XENLA	U1 SMALL NUCLEAR RIBO	1.18e+00
23	103	3.9	296	3	GDA6_WHEAT	ALPHA/BETA-GLIADIN PR	1.51e+00
24	103	3.9	367	5	P53_CHICK	CELLULAR TUMOR ANTIGE	1.51e+00
25	103	3.9	1638	1	BRM_DRONE	HOMEOIC GENE REGULAT	1.51e+00
26	102	3.9	533	6	RRXC_HUMAN	RETINOIC ACID RECEPT	1.92e+00
27	102	3.9	2477	3	FINC_RAT	FIBRONECTIN PRECURSOR	1.92e+00
28	102	3.9	464	8	VIZ_HPV56	PROBABLE L2 PROTEIN.	1.92e+00
29	102	3.9	566	5	NARQ_ECOLI	NITRATE/NITRITE SENSO	1.92e+00
30	101	3.8	141	8	YPR1_OMEFU	HYPOTHETICAL PROLINE-	2.45e+00
31	101	3.8	744	1	CA18_RABIT	COLLAGEN ALPHA 1(VIII	2.45e+00
32	101	3.8	510	3	ERM_HUMAN	ETS-RELATED PROTEIN E	2.45e+00
33	101	3.8	513	6	RRXB_HUMAN	RETINOIC ACID RECEPT	2.45e+00
34	100	3.8	267	3	EXTN_MAIZE	EXTENSIN PRECURSOR (P	3.12e+00
35	100	3.8	1748	6	POLR_ELV	RNA REPLICASE POLYPRO	3.12e+00
36	100	3.8	992	2	EBN6_EBV	EBNA-6 NUCLEAR PROTEI	3.12e+00
37	100	3.8	522	3	GAG5_HUMAN	N-ACETYLGALACTOSAMINE	3.12e+00
38	100	3.8	624	8	VP87_NPVOP	CAPSID PROTEIN P87.	3.12e+00
39	99	3.7	2274	5	MUC2_HUMAN	MUCIN 2 (INTESTINAL M	3.96e+00
40	98	3.7	1434	8	V685_HSV11	HYPOTHETICAL GENE 65	5.02e+00
41	98	3.7	455	8	VIZ2_HPV11	PROBABLE L2 PROTEIN.	5.02e+00
42	98	3.7	483	1	ATFA_HUMAN	TRANSCRIPTION FACTOR	5.02e+00
43	98	3.7	719	5	PGCA_BOVIN	CARTILAGE-SPECIFIC PR	5.02e+00
44	98	3.7	140	8	YTV6_HUMAN	HYPOTHETICAL 14 KD PR	5.02e+00
45	97	3.7	466	1	ANX7_HUMAN	ANNEXIN VII (SYNEXIN)	6.35e+00

ALIGNMENTS

RESULT	1	ID	TPO_MOUSE	STANDARD;	PRT;	356 AA.
AC	P40226;	DT	01-FEB-1995 (REL. 31, CREATED)			
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)	DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	THROMBOPOIETIN PRECURSOR (MEGAKARYOCYTE COLONY STIMULATING FACTOR)	DE	(C-MPL LIGAND) (ML) (MEGAKARYOCYTE GROWTH AND DEVELOPMENT FACTOR)			
DE	(MGDF).	OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;	OC	EUKARYOTA; RODENTIA.			
RN	[1]	RP	SEQUENCE FROM N.A.			
RM	36757807	RA	LOK S.			
RA	0087777	RA	0087777			
RA	BELL L.A.N., GRANT F.J., HEIPEL M.D., BURKHEAD S.K., KRAMER J.M.,	RA	CHING A.F.T., MATHEWS S.L., BAILEY M.C., FORSTROM J.W., BUDDLE M.M.,			
RA	OSBORNE S.G., EVANS S.J., SHEPPARD P.O., PRESNELL S.R., O'HARA P.J.,	RA	HAGEN F.S., ROTH G.J., FOSTER D.C.;			
RL	NATURE 369:565-568(1994).	CC	-/- FUNCTION: ACTS AS A CIRCULATING REGULATOR OF PLATELET NUMBERS.			
CC	MEGAKARYOCYTIC LINEAGE SPECIFIC GROWTH AND DIFFERENTIATION	CC	FACTOR.			

CC -!- SUBCELLULAR LOCATION: SECRETED.
 DR EMBL; L34169; WNTROA.
 DR PIR; S45330; S45330.
 KW GLYCOPROTEIN; HORMONE; SIGNAL.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 356 THROMBOPOEITIN.
 FT CARBOHYD 197 197 POTENTIAL.
 FT CARBOHYD 206 206 POTENTIAL.
 FT CARBOHYD 235 235 POTENTIAL.
 FT CARBOHYD 249 249 POTENTIAL.
 FT CARBOHYD 256 256 POTENTIAL.
 FT CARBOHYD 336 336 POTENTIAL.
 FT CARBOHYD 351 351 POTENTIAL.
 SQ SEQUENCE 356 AA; 37835 MW; 701294 CN;
 DB 7; Score 2482; Match 100.0%; QryMatch 93.9%; Pred. No. 0.00e+00;
 Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 meltclllaamlavaritlespvpacpdrllnkllrdshllhslscgpcdvplspv 60
 Qy 24 MELTDLILAAMLLAVARITLSSVPACDPRLLNKLLRDShLLHSLSLSCQCPDVPDPLSPV 83
 Db 61 llpavdfealgewktqtegskaqdiilgavlllegvmaargilepsclslqlglsqvrl 120
 Qy 84 LLPAVDLSGEMKQTQESKAQDIILGAVSLLLEGVMAARGILEPSCLSLLQLGSLQVRL 143
 Db 121 llgalqglgtqlpqrtaahkdpnaiflqlqllrgkvrfllllvegptlcvrtppt 180
 Qy 144 LLGALQGLGTQLPQRTAHRKDPNALFLSLQQLLRGVRFLLLLVEGPTLCVRRTPPT 203
 Db 181 avpsstqltlmkfnrtsglltnfsvrtagpqlrlgkvrflmlvggptlcvrtppt 240
 Qy 204 AVPSSTQLTLMKFNRTSGLLTNFVSVRTAGPQLRLGKVRFLMLVGGPTLCVRRTPPT 263
 Db 241 pvqisgylnrthgpnvngthglfaqtsgltleasdiqgafnkslafnlqglpsspsla 300
 Qy 264 PVQISGYLNRTHGPNVNGTHGLFAQTSGLTLEASDIQGFNKGSLAFNLQGLPSPSLA 323
 Db 301 pdghtpfpspalptthgspqllhplfpdptmpnstaphvmtphprnlsgset 356
 Qy 324 PDGHTPFPSPALPTTHGSPQLLHPLFPDPTMPNSTAPHVMTPHPRNLSGSET 379

RESULT 2
 ID TPO HUMAN STANDARD; PRT; 353 AA.
 AC P40225;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE THROMBOPOEITIN PRECURSOR (MEGAKARYOCYTE COLONY STIMULATING FACTOR)
 DE (C-MPL LIGAND) (ML) (MEGAKARYOCYTE GROWTH AND DEVELOPMENT FACTOR)
 DE (MGDF).
 GN THPO.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN (1)
 RP SEQUENCE FROM N.A.
 RM 94261202
 RA DE SAUVAGE F.J., BASS P.E., SPENCER S.D., WALLOY B.E., GURNEY A.L.,
 RA SPENCER S.A., DARGONNE W.C., HENZEL W.J., WONG S.C., KUANG W.-J.,
 RA OLES K.J., HOUTGREN B., SOLBERG L.A. JR., GOEDDEL D.V., EATON D.L.;
 RL NATURE 369:533-538 (1994).
 RN (2)

RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RM 94291201
 RA BARTLEY T.D., BOGENBERGER J., HUNT P., LI Y.S., LU H.S.,
 RA MARTIN F., CHANG M.S., SAMAL B.B., NICHOL S., BOSSELMAN R.A.;
 RL CELL 77:1117-1124 (1994).
 CC -!- FUNCTION: ACTS AS A CIRCULATING REGULATOR OF PLATELET NUMBERS.
 CC MEGAKARYOCYTIC LINEAGE SPECIFIC GROWTH AND DIFFERENTIATION
 CC FACTOR.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 DR EMBL; L33410; HSMCMPL.
 DR EMBL; U11025; HS11025.
 DR PIR; S45331; S45331.
 DR MIM; 600044; 11TH EDITION.
 KW GLYCOPROTEIN; HORMONE; SIGNAL.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 353 THROMBOPOEITIN.
 FT CARBOHYD 197 197 POTENTIAL.
 FT CARBOHYD 206 206 POTENTIAL.
 FT CARBOHYD 234 234 POTENTIAL.
 FT CARBOHYD 255 255 POTENTIAL.
 FT CARBOHYD 340 340 POTENTIAL.
 FT CARBOHYD 348 348 POTENTIAL.
 FT CONFLICT 113 113 Q -> E (IN REF. 2).
 SQ SEQUENCE 353 AA; 37822 MW; 702923 CN;

DB 7; Score 1748; Match 71.6%; QryMatch 66.1%; Pred. No. 1.52e-287;
 Matches 255; Conservative 47; Mismatches 49; Indels 5; Gaps 4;
 Db 1 meltcllvmltlltarltlssvpacpdrllnkllrdshvhlhslscgpcdvplspv 60
 Qy 24 MELTDLILAAMLLAVARITLSSVPACDPRLLNKLLRDShLLHSLSLSCQCPDVPDPLSPV 83
 Db 61 llpavdfealgewktqtegskaqdiilgavlllegvmaargilepsclslqlglsqvrl 120
 Qy 84 LLPAVDLSGEMKQTQESKAQDIILGAVSLLLEGVMAARGILEPSCLSLLQLGSLQVRL 143
 Db 121 llgalqglgtqlpqrtaahkdpnaiflqlqllrgkvrfllmlvggptlcvrtppt 180
 Qy 144 LLGALQGLGTQLPQRTAHRKDPNALFLSLQQLLRGVRFLLLLVEGPTLCVRRTPPT 203
 Db 181 avpsstqltlmkfnrtsglltnfsvrtagpqlrlgkvrflmlvggptlcvrtppt 239
 Qy 204 AVPSSTQLTLMKFNRTSGLLTNFVSVRTAGPQLRLGKVRFLMLVGGPTLCVRRTPPT 263
 Db 240 lqdiqgylnrthgpnvngthglfaqtsgltleasdiqgafnkslafnlqglpsspsla 299
 Qy 264 PVQISGYLNRTHGPNVNGTHGLFAQTSGLTLEASDIQGFNKGSLAFNLQGLPSPSLA 323
 Db 300 pdghtpfpspalptthgspqllhplfpdptmpnstaphvmtphprnlsgset 352
 Qy 324 PDGHTPFPSPALPTTHGSPQLLHPLFPDPTMPNSTAPHVMTPHPRNLSGSET 378

RESULT 3
 ID EPO MOUSE STANDARD; PRT; 192 AA.
 AC P07321;
 DT 01-APR-1988 (REL. 07, CREATED)
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE ERYTHROPOIETIN PRECURSOR.
 GN EPO.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

[illegible]

RL	BIOCHM. BIOPHYS. ACTA 1171:99-102(1992).
CC	-!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE
CC	REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF A
CC	PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.
CC	-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS
CC	AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
CC	-!- SUBCELLULAR LOCATION: SECRETED.
DR	EMBL; D10763; RNEPO.
DR	PIR; S28148; S28148.
DR	PROSITE; PS00817; EPO.
KW	ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.
FT	SIGNAL 1 26 BY SIMILARITY.
FT	CHAIN 27 192 ERYTHROPOIETIN.
FT	DISULFID 33 187 BY SIMILARITY.
FT	CARBOHYD 50 50 BY SIMILARITY.
FT	CARBOHYD 64 64 BY SIMILARITY.
FT	CARBOHYD 109 109 BY SIMILARITY.
SQ	SEQUENCE 192 AA; 21286 MW; 179242 CN;
DB	3; Score 145; Match 24.3%; QryMatch 5.5%; Pred. No. 1.46e-05;
Matches	34; Conservative 40; Mismatches 62; Indels 4; Gaps 4;
Db	11 lllslllplglpvcaprlldcsrvlerylleakeaenvmgcagprlsenitvdpdt 70
Qy	29 LLLAAMLLAVARLTLSPPACAPRMLNKLRDLSHLHSRLSQCPDVPDLSIPVLLPAV 88
Db	71 kmvfyawkmvceqavewqglleallseailqalqalqansppeslqlhdkaisglrs 130
Qy	89 DFIGEMKWTQTESKQADITLGAVALLEGGVMAARG-QLEPS-CLSSLLGSLSQVR-LL- 144
Db	131 ltsllrvlgaqdelmsppda 150
Qy	145 LGAQLGLGTQLPQGRTTA 164
RESULT	5
ID	EPO FELCA STANDARD; PRT; 188 AA.
AC	P33708;
DT	01-FEB-1994 (REL. 28, CREATED)
DT	01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT	01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE	ERYTHROPOIETIN PRECURSOR (FRAGMENT).
GN	EPO.
OS	FELIS CATUS (CAT).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC	EUTHERIA; CARNIVORA.
RN	(1)
RP	SEQUENCE FROM N.A.
RM	93372347
RA	WEN D., BOISSEL J.P.R., TRACY T.E., MULCAHY L.S., CZELUSNIAK J.,
RA	GOODMAN M., BUNN H.F.;
RA	BLOOD 82:1507-1516(1993).
CC	-!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE
CC	REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF A
CC	PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.
CC	-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS
CC	AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
CC	-!- SUBCELLULAR LOCATION: SECRETED.
DR	EMBL; L10606; FCERYTHRO.
DR	PROSITE; PS00817; EPO.
KW	ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.
FT	NON TER 1 1
FT	SIGNAL <1 22 BY SIMILARITY.
FT	CHAIN 23 188 ERYTHROPOIETIN.


```

RL GLYCOTOLOGY 1:337-346(1991).
CC -!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE
CC REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF A
CC PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.
CC -!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS
CC AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
CC -!- SUBCELLULAR LOCATION: SECRETED.
DR EMBL; X02159; HSERPG.
DR EMBL; X02157; HSERPR.
DR EMBL; M11319; HSERPA..
DR PIR; A01855; ZOHUO.
DR PIR; A25384; A25384.
DR PIR; A24744; A24744.
DR PIR; A22210; A22210.
DR MM; I33170; 11TH EDITION.
DR PROSITE; PS00817; EPO.
KW ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.
FT SIGNAL      1    27
FT CHAIN       28   193          ERYTHROPOIETIN.
FT DISULFID    34   188
FT DISULFID    56    60
FT CARBOHYD     51    51
FT CARBOHYD    65    65
FT CARBOHYD   110   110
FT CARBOHYD   153   153
FT CARBOHYD   190   193
FT PROPEP      85    85
FT CONFLICT    85    85
SQ SEQUENCE    193 AA; 21306 MW; 182311 CN;

DB 3; Score      131; Match 22.7%; QryMatch 5.0%; Pred. No. 8.68e-04;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0
Db 12 lllslslplglvlgapprlcldsrvlrvlylleakeanittgcaehcslnenitvpdt 71
Qy llll : :::: :!! :|::|| : | : : |:: || : : | : : ! : : ! :
29 LLIAAMLAVARILTSSPVAPCDPLINKLLRDLHSLHSLSQCDDVDPLSIVLPVAV 88
:: || | : : ::|| |:| : : | : : : : :
Db 72 knvfyaikrmvevgqavewggllallseavlrgqall 108
Qy 89 DFSLGEMKTKTEQSQAQDILGCAVSLLGLEGYMAARGQL 125

RESULT      8
ID EPO SHEEP STANDARD; PRT; 194 AA.
AC P33709;
DT DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE ERYTHROPOIETIN PRECURSOR.
GN EPO.
OS OVIS ARIES (SHEEP).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ARTIODACTYLIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RM 93351736
RA RA FUL., EVANS B., LIM G.B., MORITZ K., WINTOUR M.E.;
RL MOL. CELL. ENDOCRINOL. 93:107-116(1993).
CC -!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE
CC REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF A
CC PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.
CC -!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS
CC AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
CC -!- SUBCELLULAR LOCATION: SECRETED.
```

DR	EMBL; Z24681; OARPTOIA.
DR	PROSITE; PS00817; EPO.
KW	ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.
FT	SIGNAL 1 27 BY SIMILARITY.
FT	CHAIN 28 194 ERYTHROPOIETIN.
FT	DISULFID 34 189 BY SIMILARITY.
FT	DISULFID 56 60 BY SIMILARITY.
FT	CARBOHYD 51 51 POTENTIAL.
FT	CARBOHYD 65 65 POTENTIAL.
FT	CARBOHYD 110 110 POTENTIAL.
SQ	SEQUENCE 194 AA; 21335 MW; 175348 CN;
DB	3; Score 128; Match 21.6%; OryMatch 4.8%; Pred. No. 2.03e-03;
Matches	21; Conservative 35; Mismatches 41; Indels 0; Gaps 0;
Db	12 lllefllplglvgaprlldcsrvlylreaseanatmgcaegcsfsenitvdpd 71 ::: ::: ::: ::: : ::: :::
Qy	29 LLAAAMLLAVARTLSSVPACDPRLINKLRLLSHLSRLSCQCPDVPDISPVLFPV 88 :: :: ::: :
Db	72 kmfyakmtevgqqlawgqlallseafgqall 108 :: :: ::: :
Qy	89 DFIGEWMKTQTSKQAQDILGAVSLLEGVMAARGQL 125 :: :: ::: :
RESULT	9
ID	FORM MOUSE STANDARD; PRT; 1468 AA.
AC	Q05860;
DT	01-JUN-1994 (REL. 29, CREATED)
DD	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE	FORMIN (LIMB DEFORMITY PROTEIN).
GN	LD.
OS	MUS MUSCULUS (MOUSE).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC	EUTHERIA; RODENTIA.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-KIDNEY, AND TESTIS;
RM	90363291
RA	WOYCHIK R.P., MAAS R.L., ZELLER R., VOGT T.F., LEDER P.;
RL	NATURE 346:850-853(1990).
CC	-!- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
CC	AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY BE
CC	INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.
CC	-!- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
CC	LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
CC	GLAND. IT IS PRESENT THROUGHOUT THE EMBryo.
CC	-!- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN
CC	IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
CC	COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
CC	KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
CC	THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
CC	-!- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN
CC	ARE PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A
CC	VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND
CC	DIFFERENT SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
CC	EMBL; X53599; MMULD.
DR	PIR; S11515; S11515.
HSP	P199999; ICLG.
KW	NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
FT	DOMAIN 298 301 POLY-SER.
FT	DOMAIN 861 864 POLY-SER.
FT	DOMAIN 869 970 PRO-RICH.

FT	DOMAIN	977	901	POLY-SER.
5Q	SEQUENCE	1468 AA;	163809 MW;	11493196 CN;
DB	3;	Score	117;	Match 40.0%; QryMatch 4.4%; Pred. No. 4,14e-02;
	Matches	20;	Conservative	12; Mismatches 16; Indels 2; Gaps 2;
Db	884	ipppplpp-glqllppappipppcvspppppp-pppctpvppsdgppp	931	
		: : : : : : : : : : : : : : :		
Qy	316	lppspslapdghtppfppspalptthgspqllhplffpdpsttmnstaphp	365	
		: : : : : : : : : : : : : : : :		
RESULT	10			
ID	FOR4 MOUSE	STANDARD;	PRT;	1206 AA.
AC	Q05859;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	FORMIN 4 (LIMB DEFORMITY PROTEIN).			
GN	LD.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
NC	EUTHERIA; RODENTIA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=EMBRYO;			
RM	92112033			
RA	GRUBBY-JACKSON L., KUO A., LEDER P.;			
RL	GENES DEV. 6:29-37(1992).			
CC	-/- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB			
CC	AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR			
CC	BE INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.			
CC	-/- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT			
CC	HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.			
CC	-/- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL			
CC	ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING			
CC	LIMB BUD.			
CC	-/- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE			
CC	PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN			
CC	SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE			
CC	TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS			
CC	DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR			
CC	DETERMINED.			
DR	EMBL; X62379; WMIOR.			
DR	PIR; S24407; S24407.			
DR	HSP; P19999; ICLG.			
FW	NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.			
KT	DOMAIN 635 638 POLY-SER.			
FT	DOMAIN 644 744 PRO-RICH.			
FT	DOMAIN 751 755 POLY-SER.			
5Q	SEQUENCE	1206 AA;	133464 MW;	7742033 CN;
DB	3;	Score	117;	Match 40.0%; QryMatch 4.4%; Pred. No. 4,14e-02;
	Matches	20;	Conservative	12; Mismatches 16; Indels 2; Gaps 2;
Db	658	ipppplpp-glqllppappipppcvspppppp-pppctpvppsdgppp	705	
		: : : : : : : : : : : : : : :		
Qy	316	lppspslapdghtppfppspalptthgspqllhplffpdpsttmnstaphp	365	
		: :~ :~ :~ :~ :~ :~ :~ :~ :~ :~ :~ :~		
RESULT	11			
ID	EPO MACFA	STANDARD;	PRT;	192 AA.
AC	P07865;			
DT	01-AUG-1988 (REL. 08, CREATED)			
DT	01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)			

```
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
GN    ERYTHROPOIETIN PRECURSOR.
OS MACACA FASCICULARIS (CRAB EATING MACAQUE) (CYNOMOLGUS MONKEY).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CN EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 87055236
RL LIN F.-K., LIN C.-H., LAI P.-H., BROWNE J.K., EGRIE J.C., SMALLING R.,
RA FOX G.M., CHEN K.K., CASTRO M., SUGGS S.;
RG GENIE 44:201-209(1986).
CC -!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE
CC REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF A
CC PHYSIOLOGICAL LEVEL OF CIRCUATING ERYTHROCYTE MASS.
CC -!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS
CC AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
CC -!- SUBCELLULAR LOCATION: SECRETED.
DR EMBL; M18189; MFEP.
DR EMBL; M18188; MFEPD.
DR PIR; JQ0173; JQ0173.
DR PROSITE; PS00817; EPO.
KW ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.
FT SIGNAL      1 27       BY SIMILARITY.
FT CHAIN       28 192     ERYTHROPOIETIN.
FT DISULFID    34 187     BY SIMILARITY.
FT DISULFID    56 60      BY SIMILARITY.
FT CARBOHYD    51 51      BY SIMILARITY.
FT CARBOHYD    65 65      BY SIMILARITY.
FT CARBOHYD   110 110     BY SIMILARITY.
FT CARBOHYD   152 152     BY SIMILARITY.
SQ SEQUENCE    192 AA; 21113 MW; 175216 CN;

DB 3; Score 114; Match 22.0%; QryMatch 4.3%; Pred. No. 9.17e-02;
Matches 22; Conservative 32; Mismatches 46; Indels 0; Gaps 0;

Db 12 lllslvslplqvpagprlrcdsrvlyrlleakeavntmcscsechenltvdpd 71
||| : ||:: ::||:||||:: |:: : ||: ::| :|
Qy 29 LLTAAALLAVARTLTSSVPACDPRLINKLRDLSHLLHSRLSQCPDVPLSPVLPAPV 88

Db 72 kmvfayakrmvggqvawegqlallseavlrgqavlans 111
:: ||         |:|:: ::||| l:l: ::| |
Qy 89 DFSIGEMKTQTQSQAQDILGVASLLLEGVMARGQLEPS 128

RESULT 12
ID RUIC HUMAN STANDARD; PRT; 159 AA.
AC P09234;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE U1 SMALL NUCLEAR RIBONUCLEOPROTEIN C.
GN SNRPC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CN EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 88335591
RL SILLECKENS P.T.G., BEIJER R.P., HABETS W.J., VAN VENROOIJ W.J.;
NL NUCLEIC ACIDS RES. 16:8307-8321(1968).
RN [2]
RP SEQUENCE OF 25-131 FROM N.A.
RM 88088826
```

RA YAMAMOTO K., MIURA H., MOROI Y., YOSHINOYA S., GOTO M., NISHIOKA K.,
 RA MIYAMOTO T.;
 * RL J. IMMUNOL. 140:311-317(1988).
 - CC -!- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U1.
 - CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 - DR EMBL; X12517; HSURNPC.
 - DR EMBL; M18465; HSSNRPA.
 DR PIR; S01387; S01387.
 KW NUCLEAR PROTEIN; RNA-BINDING; RIBONUCLEOPROTEIN; ZINC-FINGER.
 FT 2N FING 6 30 POTENTIAL.
 FT CONFLICT 25 27 CSG -> RSR (IN REF. 2).
 FT CONFLICT 94 98 APHMG -> TPIW (IN REF. 2).
 FT CONFLICT 101 101 P -> S (IN REF. 2).
 FT CONFLICT 129 131 HMP -> ICQ (IN REF. 2).
 SQ SEQUENCE 159 AA; 17394 MW; 144050 CN;

DB 6; Score 113; Match 31.4%; QryMatch 4.3%; Pred. No. 1.19e-01;
 Matches 22; Conservative 18; Mismatches 27; Indels 3; Gaps 3;

Db 51 dktaafqgkipptfaappagamiapppslpgpprpgmmaphmgagppmmgagppp 110
 QY I : : I I : : : : I I : : : : I I : : : : I I : : : : I I : : : :
 297 DISGAFNKGSLAFNIQGLSPSLAPDGT-PFPPSPAL-PTTH-GSPQLHLFPDP 353
 Db 111 pgmmpvgpap 120
 QY : : I I : :
 354 STTMENSTAP 363

RESULT 13

ID EXTN TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
 DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HRCNPT3.
 OS NICOTIANA TABACUM (COMMON TOBACCO).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 OC SOLANALES; SOLANACEAE.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=CV. XANTHI; TISSUE=LEAF;
 RM 90128263
 RA KELLER B., LAMB C.J.;
 RL GENES DEV. 3:1639-1646(1989).
 CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 CC MAIN ROOT.
 CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYCOSYLATED.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 DR EMBL; X13885; NTEXT.
 DR PIR; S06733; S06733.
 DR HSP; P19999; ICLG.
 KW REPEAT; CELL WALL; GLYCOPROTEIN; SIGNAL; STRUCTURAL PROTEIN;
 KW HYDROXYLATION.
 FT SIGNAL 1 ?
 FT CHAIN 2 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 242 2.

FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 SQ SEQUENCE 620 AA; 65406 MW; 2614500 CN;
 DB 3; Score 112; Match 33.3%; QryMatch 4.2%; Pred. No. 1.55e-01;
 Matches 21; Conservative 16; Mismatches 23; Indels 3; Gaps 3;
 Db 369 lpppppspp-ppsfspppptyeqppppppaypplpapptyppppptypppptyacpp 427
 QY I I I I : : I I I I : : I I I I : : I I I I : : I I I I : : I I I I : :
 316 LPSPSLAPDGTTP-PSPALPTTHGSPQLHLFPDPST-TMENSTAPHVWYMPHR 373
 Db 428 pfp 430
 QY I :
 374 NLS 376

RESULT 14

ID GDF1 HUMAN STANDARD; PRT; 372 AA.
 AC P27539;
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE EMBRYONIC GROWTH FACTOR GDF-1 PRECURSOR.
 GN GDF1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 91239545
 RA LEE S.J.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:4250-4254(1991).
 CC -!- FUNCTION: GDF-1 MAY MEDIATE CELL DIFFERENTIATION EVENTS DURING
 CC EMBRYONIC DEVELOPMENT.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE BRAIN.
 CC -!- SIMILARITY: TO OTHER GROWTH FACTORS OF THE TGF-BETA FAMILY.
 DR EMBL; M62302; HSGDF1.
 DR PIR; C39364; C39364.
 DR PROSITE; PS00250; TGF BETA.
 KW SIGNAL; GROWTH FACTOR; CYTOKINE; GLYCOPROTEIN.
 FT SIGNAL 1 ? POTENTIAL.
 FT PROPEP ? 253 POTENTIAL.
 FT CHAIN 254 372 EMBRYONIC GROWTH FACTOR GDF-1.
 FT DISULFID 267 337 BY SIMILARITY.
 FT DISULFID 296 369 BY SIMILARITY.
 FT DISULFID 300 371 BY SIMILARITY.
 FT DISULFID 336 336 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 206 206 POTENTIAL.
 SQ SEQUENCE 372 AA; 39502 MW; 618754 CN;

DB 3; Score 108; Match 42.9%; QryMatch 4.1%; Pred. No. 4.32e-01;
 Matches 21; Conservative 7; Mismatches 19; Indels 2; Gaps 2;

Db 14 llllalllpplp-lt-rapvpqpaaallqalrdepgagaprlrppvpv 61
 QY I I I I : : I I : : I I : : I I : : I I : : I I : : I I : :
 29 LILAMLAIVARLTLSSVPAPACDPLINKL-LRSHLLHSRLSQCPDV 76

RESULT 15

ID GDA3 WHEAT STANDARD; PRT; 282 AA.
 AC P04723;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)

```

DE  ALPHA/BETA-GLIADIN PRECURSOR (PROLAMIN) (CLASS A-III).
OS  TRITICUM AESTIVUM (WHEAT).
*  OC  EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
*  OC  CYPERALES; GRAMINEAE.
*  RN  (1)
*  RP  SEQUENCE FROM N.A.
RA  85234522
RL  OKITA T.W., CHEESBROUGH V., REEVES C.D.;
    J. BIOL. CHEM. 260:8203-8213(1985).
CC  -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC  -!- THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5 HOMOLOGY CLASSES.
CC  SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO SINGLE BASE
CC  SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS WITHIN OR NEAR
CC  DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF THE GENE
CC  FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
DR  EMBL; M11076; TAGLIABH.
DR  PIR; E22364; E22364.
KW  SEED STORAGE PROTEIN; REPEAT; SIGNAL; MULTIGENE FAMILY.
FT  SIGNAL 1 20
FT  CHAIN 21 282 ALPHA/BETA-GLIADIN A-III.
SQ  SEQUENCE 282 AA; 32236 MW; 403610 CN;

DB 3; Score 107; Match 32.8%; QryMatch 4.0%; Pred. No. 5.57e-01;
Matches 22; Conservative 13; Mismatches 27; Indels 5; Gaps 5;

Db 55 qefppqpy-ph-qpfpsqpyppqpfpqql-p-ypqtqpfppqpyppqpqpypp 110
   | :|| : | : ||| : | : ||| | : | : ||| | : |
Qy 313 QGGLPPSPSLAPDGHPTFFPSPALPTTHGSPQLHPLFPDPSTTMENSTAFHPVTMYHP 372

Db 111 qpqsqq 117
   : :||:
Qy 373 RN-LSQE 378

```

Search completed: Wed Sep 27 11:39:53 1995
 Job time : 22 secs.


```

c 22 20 1.9 427 57 T95155 ye39d06.r1 Homo sapie 2.41e-01
c 23 20 1.9 488 52 T77076 yd72h11.s1 Homo sapie 2.41e-01
c 24 20 1.9 302 32 RIC00791A Rice cDNA, partial se 2.41e-01
c 25 20 1.9 488 52 T75477 yd63b01.s1 Homo sapie 2.41e-01
c 26 20 1.9 236 36 T07633 EST05523 Homo sapiens 2.41e-01
c 27 20 1.9 359 3 CELK02603F C.elegans cDNA clone 2.41e-01
c 28 20 1.9 332 44 T46575 9838 Arabidopsis thal 2.41e-01
c 29 20 1.9 307 39 T27044 NIBT203H05R Homo sapi 2.41e-01
c 30 20 1.9 475 48 T62825 yf70b12.s1 Homo sapie 2.41e-01
c 31 20 1.9 414 18 R14376 yf82a06.r1 Homo sapie 2.41e-01
c 32 20 1.9 268 46 T51043 yb71b11.r1 Homo sapie 2.41e-01
c 33 20 1.9 361 18 R13977 yf68f04.r1 Homo sapie 2.41e-01
c 34 20 1.9 331 5 HSC098051 H. sapiens partial cD 2.41e-01
c 35 20 1.9 326 57 T95459 ye46g06.r1 Homo sapie 2.41e-01
c 36 20 1.9 325 13 M62087 EST00145 Homo sapiens 2.41e-01
c 37 20 1.9 426 58 T97685 ye54a02.s1 Homo sapie 2.41e-01
c 38 20 1.9 443 58 T99311 ye63b11.r1 Homo sapie 2.41e-01
c 39 20 1.9 480 56 T90434 ye44b08.s1 Homo sapie 2.41e-01
c 40 20 1.9 323 29 R52051 yj71e09.s1 Homo sapie 2.41e-01
c 41 20 1.9 436 51 T71049 yc50g08.s1 Homo sapie 2.41e-01
c 42 20 1.9 562 44 T44094 7357 Arabidopsis thal 2.41e-01
c 43 20 1.9 290 38 T19301 e04006s Homo sapiens 2.41e-01
c 44 20 1.9 343 41 T31099 EST27367 Homo sapiens 2.41e-01
c 45 20 1.9 343 41 T31099 EST27367 Homo sapiens 2.41e-01

```

ALIGNMENTS

```

1
RESULT 1
LOCUS HSC1KF112 282 bp RNA EST 28-JAN-1995
DEFINITION H. sapiens partial cDNA sequence; clone c-1kf11.
ACCESSION F03047
KEYWORDS partial cDNA sequence; transcribed sequence fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa/Eumycota group;
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catasthini; Hominidae; Homo.
1 (bases 1 to 282)
REFERENCE
AUTHORS Genexpress.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) to the EMBL/GenBank/DBJ databases.
Genethon, B.P. 60, 91002 Evry Cedex France and Genetique
Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801
Villejuif Cedex France.E-mail: genexpress@genethon.fr
2 (bases 1 to 282)
REFERENCE
AUTHORS Genexpress.
TITLE The Genexpress cDNA program
JOURNAL Unpublished
3 (bases 1 to 282)
AUTHORS Auffray,C., Behar,G., Bois,F., Boucher,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: Integrated molecular analysis of the human genome and its
expression
C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)
Cloning method: total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the lsmid BA
vector;
Sequencing method: single read, full automatic;
Primer: (-21)/M13_universal;

```

cDNA sequence complementary to mRNA (3'end)
Stretch removed: removed at sequence 5'end
Normalization method: Bento Soares, P.N.A.S in press;
Genexpress_library_idt: C;
Genexpress_sequence_idt: aic-1kf11.

NCBI gi: 546604
FEATURES
source Location/Qualifiers
1..282
/organism="Homo sapiens"
/clone_lib="normalized infant brain cDNA from B.Souares,
Psychiatry Dept. Columbia University USA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
/isolate="muscular atrophy patient"

BASE COUNT 84 a 93 c 48 g 53 t 4 others
ORIGIN
DB 7; Score 24; Match 66.7%; QryMatch 2.3%; Pred. No. 1.14e-05;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Db 86 ctccggagacacccaccccgaccctgcgccacccctccctccgattccacatcc 145
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 825 CTCAGGAACATCAGACACAGGCTCCTGCCACCAAGCTCCAGCTGATTCCTC 884
Db 146 ctgagccatcc 157
| ||||| |||
Qy 885 CCCAACCCATCC 896

```

2
RESULT 2
LOCUS HSC1LH062 224 bp RNA EST 28-JAN-1995
DEFINITION H. sapiens partial cDNA sequence; clone c-1lh06.
ACCESSION F03088
KEYWORDS partial cDNA sequence; transcribed sequence fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa/Eumycota group;
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catasthini; Hominidae; Homo.
1 (bases 1 to 224)
REFERENCE
AUTHORS Genexpress.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) to the EMBL/GenBank/DBJ databases.
Genethon, B.P. 60, 91002 Evry Cedex France and Genetique
Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801
Villejuif Cedex France.E-mail: genexpress@genethon.fr
2 (bases 1 to 224)
REFERENCE
AUTHORS Genexpress.
TITLE The Genexpress cDNA program
JOURNAL Unpublished
3 (bases 1 to 224)
AUTHORS Auffray,C., Behar,G., Bois,F., Boucher,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: Integrated molecular analysis of the human genome and its
expression
C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)
Cloning method: total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the lsmid BA

```

```

/organism="Oryza sativa"
/strain="Nipponbare"

```

Matches 41;

DB 25; Score 23; Match 65.3%; QryMatch 2.2%; Pred. No. 1.60e-04; Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

[illegible]

ORGANISM Homo sapiens
Eucaryotae: Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 434)
Hillier, L., Clark, N., Dubouche, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasaks, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE
JOURNAL

Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Source: IMAGE Consortium, LINL
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES	Location/Qualifiers
NCBI gi.	1..434
source	

```

/organism="Homo sapiens"
/clone="67051"

```

BASE COUNT	124 a	86 c	102 g	117 t	5 others
------------	-------	------	-------	-------	----------

DB 50; Score 21; Match 76.9%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 112 agcctctcttaagctctaaccacacccaggcacagaat 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT	9	R13972	344 bp	mRNA	EST	12-APR-1995
LOCUS						

DETERMINATION
 R13972
 ACCESSION
 R13972
 KEYWORDS
 EST.
 J130000.11 HOMO SAPIENS CHINA C1000 27404 3

SOURCE human clone=27404 library=Soares infant brain INIB vector=Lafmid BA

strand cDNA was primed with a Not I - oligo(dT) primer [5' AAC TGAAGAATTCGGCGCGCCAGCAATTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 344)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlff, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Zannis-Hadjopoulos, D.

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 361
Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 786970
Location/Qualifiers
1..475
/organism="Homo sapiens"
/clone="J34193"
/note="human"

BASE COUNT 127 a 113 c 122 g 108 t 5 others
ORIGIN

DB 23; Score 21; Match 68.0%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 303 agaagccagaaggacctgccacgaaggngccccaccacgaagcaggca 352
|| || | ||| || | || | || | || | || | || | || |
Cp 365 AGGAGCAGCGACTGTCCGAAGAAGCTGCCCGCAGGAGCATGACGCA 316

RESULT 11
LOCUS R07547 388 bp mRNA EST 05-APR-1995
DEFINITION ye97e06.r1 Homo sapiens cDNA clone 125698 5'.
ACCESSION R07547
KEYWORDS EST.
SOURCE human clone=125698 library=Soares fetal liver spleen INFLS
vector=p773D (Pharmacia) with a modified polylinker host=DHI0B
(ampicillin resistant) primer=M13RP1 ReiteI=Pac I Rsite2=Eco RI
Liver and spleen from a 20 week-post conception male fetus. 1st
strand cDNA was primed with a Pac I - oligo(dT) primer [5'
AACTGGAGAAATTAATTAAAGCTTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Pac I and Eco RI sites of the modified p773
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini, Homidae; Homo.
1 (bases 1 to 388)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

REFERENCE AUTHORS
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 303
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the

/dev stage="3 months old"
/isolates="muscular atrophy patient"
/tissue type="total brain"
/clone_lib="normalized infant brain cDNA"
/sex="Female"

BASE COUNT 76 a 68 c 65 g 74 t 2 others
ORIGIN

DB 6; Score 21; Match 92.0%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 141 ctctttcccaaccatctctctac 165
||||| ||||| ||||| ||||| |||||

Qy 878 CTCCTTCCCMACCCATCTCTCTAC 902

RESULT 14
LOCUS R1CR0043A 230 bp mRNA EST 26-MAY-1995
DEFINITION Rice cDNA, partial sequence (R0043_1A).
ACCESSION D23738
KEYWORDS EST (expressed sequence tag).
SOURCE Oryza sativa (strain Nipponbare,) Seedling Root cDNA to mRNA.
ORGANISM Oryza sativa
Eukaryotes; mitochondrial eukaryotes; Chlorophyta/Embryophyta group; Charophyta/Embryophyta group; Embryophyta; Magnoliophyta; Liliopsida; Commelinidae; Poales; Poaceae; Oryza.
1 (bases 1 to 230)
Minobe, Y. and Sasaki, T.
Rice cDNA from root
Unpublished (1993)
JOURNAL Submitted (2-NOV-1993) to DBJ by:
COMMENT Yuzo Minobe
Dept. Rice Genome Research Program
National Institute of Agrobiological Resources
Kannonndai 2-1-2
Tsukuba, Ibaraki
Japan
Phone: 0298-38-7441
Fax: 0298-38-7468
PROJECT = 'RGP'.

NCBI gi: 427605
FEATURES Location/Qualifiers
source 1..230
/organism="Oryza sativa"
/strain="Nipponbare"
/dev stage="Seedling"
/sequenced_mol="cDNA to mRNA"
/tissue_type="Root"
BASE COUNT 67 a 73 c 34 g 51 t 5 others
ORIGIN

DB 33; Score 21; Match 74.4%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 53 cccaccttcncaacgcacgcacgcgcgcgcgcctcttnt 91
||| ||| ||||| ||| ||| ||||| |||

Qy 978 CCCTCTGCTCCACAGCCGCCACCTCAGCAGCTCTCT 1016

RESULT 15
LOCUS T87100 551 bp mRNA EST 17-MAR-1995
DEFINITION -yd88e09.r1 Homo sapiens cDNA clone 115336 5' similar to gb:X54363 (HUMAN);.

ACCESSION T87100
KEYWORDS EST.
SOURCE human clones-115336 library=Soares fetal liver spleen INFLS vector=pf7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13RPI Rsite=Pac I Reite2=Eco RI Liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' ACTCGAAGATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p7T3 I vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 551)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 354
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 715452
FEATURES Location/Qualifiers
source 1..551
/organism="Homo sapiens"
/clone="115336"
/note="human"

BASE COUNT 108 a 137 c 151 g 152 t 3 others
ORIGIN

DB 55; Score 21; Match 80.6%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 451 ggggtcctctgggtcctctgctgaggtacag 481
||||| ||| ||||| ||||| ||||| |||||

Cp 815 GGGGCTCTAGGCTCTGCTGAGGCTCCAG 785

Search completed: Wed Sep 27 23:48:17 1995
Job time : 252 secs.

WAPSELA
 (TH)

Release 2.1D John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
 Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Sep 28 00:17:09 1995; MspPar time 61.85 Seconds
 876.152 Million cell updates/sec

Tabular output not generated.

Title: >US-08-252-491-18
 Description: (1:1062) from US08252491.seq
 Perfect Score: 1062
 N.A. Sequence: 1 ATGGAGCTGACTGATTCCT.....ATCTGTCAGAGAGGTAA 1062
 Comp: TACCTCGACTGACTTAACCA.....TAGACAGAGCTCTCCCAT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 61539 seqs, 25515148 bases x 2

Database: n-genseq

- 1 n-gen1
- 2 n-gen2
- 3 n-gen3
- 4 n-gen4
- 5 n-gen5
- 6 n-gen6
- 7 n-gen7
- 8 n-gen8
- 9 n-gen9
- 10 n-gen10
- 11 n-gen11

Statistics: Mean 9.244; Variance 5.815; scale 1.590

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	84	7.9	1047	2	Q10572 Human Natriuretic Pep	4.67e-35
2	61	5.7	1047	2	Q10572 Human Natriuretic Pep	4.60e-21
3	44	4.1	91	9	Q51746 Oligonucleotide probe	3.04e-11
4	42	4.0	204	1	N81164 Base substituted E.co	3.91e-10

c	5	40	3.8	91	9	Q51746	Oligonucleotide probe	4.88e-09
c	6	40	3.8	204	1	N81164	Base substituted E.co	4.88e-09
c	7	29	2.7	565	6	Q35072	HCV envelope region n	2.60e-03
c	8	28	2.6	501	3	N50030	Sequence encoding new	7.98e-03
c	9	27	2.5	10596	9	Q51731	Plasmid pCiEBON for	2.41e-02
c	10	26	2.4	501	3	N50023	Sequence encoding new	7.17e-02
c	11	25	2.4	3871	2	N71302	HSV-1 gB and surround	2.09e-01
c	12	25	2.4	501	3	N50031	Sequence encoding new	2.09e-01
c	13	25	2.4	1956	10	Q58731	Encodes secretory alk	2.09e-01
c	14	24	2.3	565	6	Q35072	HCV envelope region n	5.98e-01
c	15	24	2.3	501	3	N50029	Sequence encoding new	5.98e-01
c	16	24	2.3	501	3	N50027	Sequence encoding new	5.98e-01
c	17	24	2.3	501	3	N50032	Sequence encoding new	5.98e-01
c	18	24	2.3	501	3	N50026	Sequence encoding new	5.98e-01
c	19	24	2.3	501	3	N50024	Sequence encoding new	5.98e-01
c	20	23	2.2	498	3	N50034	Sequence encoding new	1.67e+00
c	21	23	2.2	17041	3	Q21065	Genomic DNA of human	1.67e+00
c	22	23	2.2	501	3	N50028	Sequence encoding new	1.67e+00
c	23	22	2.1	920	5	Q32366	MAGE-10 genomic DNA.	4.57e+00
c	24	22	2.1	501	3	N50033	Sequence encoding new	4.57e+00
c	25	22	2.1	501	3	N50025	Sequence encoding new	4.57e+00
c	26	21	2.0	943	5	Q32356	MAGE-31 gene.	1.22e+01
c	27	21	2.0	1914	7	Q41701	Rat serotonin clone S	1.22e+01
c	28	21	2.0	2108	7	Q41702	Rat serotonin clone S	1.22e+01
c	29	20	1.9	2472	7	Q43967	Pokeweed antiviral pr	3.16e+01
c	30	20	1.9	3437	3	Q14937	Abelson Related Gene,	3.16e+01
c	31	20	1.9	7876	10	Q58707	E. coli multiple anti	3.16e+01
c	32	20	1.9	501	3	N50023	Sequence encoding new	3.16e+01
c	33	20	1.9	4481	1	N90958	Sequence encoding ger	3.16e+01
c	34	20	1.9	15672	2	Q10613	Rianodin receptor gen	3.16e+01
c	35	20	1.9	4487	3	Q20514	Encodes germ cell alk	3.16e+01
c	36	20	1.9	1409	3	N50474	Sequence of brain spe	3.16e+01
c	37	20	1.9	4258	3	Q22439	DNA of hgEco-1, encod	3.16e+01
c	38	20	1.9	12752	8	Q47380	Sequence which corres	3.16e+01
c	39	20	1.9	1237	2	N70314	Sequence encoding hum	3.16e+01
c	40	20	1.9	541	3	N50199	Sequence of cDNA clon	3.16e+01
c	41	20	1.9	3230	4	Q25595	Human Gnt I cDNA clon	3.16e+01
c	42	20	1.9	3871	2	N71302	HSV-1 gB and surround	3.16e+01
c	43	20	1.9	1265	6	Q36503	Sequence of zucchini	3.16e+01
c	44	20	1.9	325	8	Q59215	Human brain Expressed	3.16e+01
c	45	19	1.8	2226	5	Q32360	MAGE-5 cDNA.	7.93e+01

ALIGNMENTS

RESULT 1
 ID Q10572 standard; DNA; 1047 BP.
 AC Q10572;
 DT 09-APR-1991 (first entry)
 DE Human Natriuretic Peptide Receptor B.
 KW NPB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
 KW hyperaldosteronism; glaucoma; guanyl cyclase.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= signal sequence
 FT Protein 12
 FT /label= mature NPB
 FT Domain 23..455
 FT /label= extracellular domain
 FT /note= *binds natriuretic peptides A, B and C*
 FT Domain 456..456
 FT /label= transmembrane domain
 FT Domain 479..1047

FT /label= cytoplasmic domain
 FT /note= "GC and protein kinase activity"
 FT Modified -site 24..26
 FT /label= N-glycos_site
 FT Modified -site 35..37
 FT /label= N-glycos_site
 FT Modified -site 161..163
 FT /label= N-glycos_site
 FT Modified -site 195..197
 FT /label= N-glycos_site
 FT Modified -site 244..246
 FT /label= N-glycos_site
 FT Modified -site 277..279
 FT /label= N-glycos_site
 FT Modified -site 349..351
 FT /label= N-glycos_site
 FT Modified -site 600..602
 FT /label= N-glycos_site
 PN W09100292-A.
 PD 10-JAN-1991.
 PF 22-JUN-1989; U03586.
 PR 23-JUN-1989; US-370673.
 PA (GETH) GENENTECH INC.
 PI Chang M, Goeddel D, Lowe D;
 DR WP1; 91-036711/05.
 DR N-PSDB; Q10324.
 PT Natriuretic protein receptor B - for diagnosis and treatment of
 PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
 PS Claim 3; Fig 1; 49pp; English.
 CC The sequence was derived from the DNA encoding natriuretic peptide
 CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein
 CC kinase activity. The DNA can be inserted into expression vectors
 CC for the prodn. of the protein, opt. after being mutated to produce
 CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
 CC 114,952). The protein (or variants) can be used in treatment of
 CC natriuretic peptide disorders, and also to isolate peptides using
 CC affinity chromatography. Antibodies with affinity for NPRB can
 CC also be prep'd.
 SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;
 DB 2; Score 84; Match 9.3%; QryMatch 7.9%; Pred. No. 4.67e-35;
 Matches 89; Conservative 285; Mismatches 571; Indels 17; Gaps 17;
 Db 76 synynansavdknkyhdndnngvcvynaasvarnashwrnnnnntagavaasgnakndh 135
 Cp 1020 GTTTAGAGAGGGCTGGTAGGGGTGGCGCTGGACAGAGGGGTGAGGAGCGGGGTG 961
 Db 136 yrtvrtqnsankngnvvntghnwtaraamnyndartddrnhyntngvnnanngsn 195
 Cp 960 GAGCTGACACAGGGGTGGGAGAGGTGGTGGAGAGGAGGACCGGTATATCTGTCCAGT 901
 Db 196 nevnhvnyrnngnnnathnnrangrnyvncggnnnnnnnnnnnnnnnnnnnngd-yvyny 254
 Cp 900 AGGAGAGTGGTGGGAGGAGGAATATCCAGGCTGGAGGTTGGGTGGCAGGAGCGCTGT 841
 Db 255 ndvngnsnagntkatgrndntrnnanraanntvntyrnnnnnnnnnnnnnnnnnn 314
 Cp 840 GTCTGATGTTCTGAGGAATCTCCGGGGCTCTTAGGGTCTCGCTGAGGGTCCAGGAAA 781
 Db 315 nrarn-dhvyngnsnnnnnagcydgnnyanvnnntnnngtrndgnrnnvnmngry 373
 Cp 780 GAGTCCACGAGTCCATTCAAGAGTGGTGTATCTGTTCAGGTATCCGGGGATTTGGTC 721
 Db 374 hqvtgtnvnmkndkndrntdnvnmwngdhndgdnnaabyganknnwrtgrnnnnwkgann 433

Cp 720 CAG-GGACCTGGAGGTTTGGTTTCAG-CAGACGAGCAATCT-TGGCTCTGAATCCCTGCTG 664
 Db 434 edhnnncandhdnsedktnnetnnavngtgnntnmngvssnnnnrknmmnknnasmwr 493
 Cp 663 CCACCTTCAGAGCCAGACGAGGAGTAGTTCTGGCTGAGGAGTGAAAGTTTGTCTCCACAA 604
 Db 494 nrwnnnmgnsrnyhkgasrntnsrsgvsnmtahgkyknnnnantghnkgvnnvankhv 553
 Cp 603 TCACAGAGTCTCTGTTGGGAGCTGCTTCACTGTGAGGACTAGAGAGTTCTGCTGGGAC 544
 Db 554 nk-krrnntnrnnnnnkhmrdrvnnhnrungacndnnnnnncvntnrcnrgsnndhnnnds 612
 Cp 543 AGCTGTGGTGGGGGGGGGGGGCTGAC-GCAGAGGGTGGACCCCT-CCTA-CAAGCATCAG 487
 Db 613 nmndmmnrysnndnvkgmannhnsnshgsksnscvsvdtrnvknntdygnaenreta 672
 Cp 486 GAAACGACACTT-TCCTCGGAGCAGGTGTGGAAGCTCAGGAGGATGAGGCAAGTG- 430
 Db 673 nmddnnanyakkntannnnnngnnnttgmnaadvysngnnnnnnnnnnnnnnnnngndns 732
 Cp 429 CTTGTGAGCTGTGCTGCTGCTGCTGCTGAG-GAAGCTGGCTTCCAAGGAGGCTCTGCAGGG 371
 Db 733 nknnvkvrvngnrnrynrsndrtnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 792
 Cp 370 CCCAGGAGGAGGAGGAGGAGCTGTCCAGAAAGCTGCCCGAGGAGGATGAGGCAAGTG- 312
 Db 793 ggtennndnn 852
 Cp 311 GGTCCAGAGTGTCCCGCGCTGCCATCAC-TCCCTCCACAGAGAGGCTCACTGTCTCCAG 253
 Db 853 nanandsvtvnsdngntansanstnmvntnnndhntcndannndndvkvntngday 912
 Cp 252 AATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 193
 Db 913 mvveqngnrgnrhannmarannandavsnrnrhnrhndnnrnrn-gvhtgnv-eagvv 970
 Cp 192 AGCAGGAGCAGCAGCAGGCTGTAGGCAAGGGTGAACCTCTGGCAGCTGGCTCAGTCTGCT 133
 Db 971 gnkmyrcnngdvtntasrnsngnanknhvssttkddandnngcnmmnnrvgdmkqkqk 1030
 Cp 132 CTGAGGAGCATGGGAGCTCAGCAGGAGCTTACTGAGGACTCGGAGGCTCAGAGCAGGAGG 73
 Db 1031 mr 1032
 Cp 72 AG 71
 RESULT 2
 ID Q10572 standard; DNA; 1047 BP.
 AC Q10572;
 DT 09-APR-1991 (first entry)
 DE Human Natriuretic Peptide Receptor B.
 KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
 KW hyperaldosteronism; glaucoma; guanylyl cyclase.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= signal sequence
 FT Protein 12
 FT /label= mature NPRB
 FT Domain 23..455
 FT /label= extracellular domain
 FT /note= "binds natriuretic peptides A, B and C]"


```

12 svhsyvvvvhshhswhvhhvhhsvvvvhhvhhvhhvhhvhhvsvvc 61
      :::::::::: :::::::::: ::::::::::::::::::::::|
960 CCACCCCGCTGCTTCTGAGCCCTTCTGCTCCAGCCGCCACCTTACGAGCC 1009

RESULT
4      ID N81164 standard; DNA; 204 BP.
AC      N81164;
08-NOV-1990 (first entry)
Base substituted E.coli beta-galactosidase alpha-fragment.
E.coli beta galactosidase alpha-fragment; base substitutions; ss.
Escherichia coli.
Key      Location/Qualifiers
FF      misc feature 19..69
FF      /tag= a
FF      /function= multiple cloning site
FF      primer bind 187..204
FF      /tag= b
FF      /tag= _b
FF      EP-285123-A.
FF      05-MAY-1988.
FF      30-MAR-1988; 105163.
FF      03-APR-1987; US-034819.
FF      (SUSO) SUOMEN SOKERI OY.
FF      Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
FF      WPI-88-279927/40.
FF      Introducing random point mutations into nucleic acids -
FF      by prepn of single stranded template, annealing a primer, elongation,
FF      misincorporation, completion of molecules and screening.
FF      Disclosure; p; English.
FF      Random point mutations were introduced into the alpha fragment of
FF      E.coli beta-galactosidase. The wild type sequence was obtained as a
FF      single stranded template and an oligonucleotide was hybridised to
FF      it to generate a popn of DNA molecules which terminate at all
FF      possible nucleotide positions within a specified region. The
FF      variable 3' ends generated in this way are used as primers for
FF      reverse transcriptase. Nucleotides are misincorporated by the
FF      transcriptase and the molecules are completed to forms that can be
FF      amplified and then expressed in a suitable host-vector system.
FF      The sequence covers all 176 diffit base substitutions, most of which
FF      occurred singularly in any given mutant.
FF      See also P80575.
FF      Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

```

[illegible]

Db	189 ch 190
	:
Dy	1004 CC 1005

RESULT 5
IID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A

Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
 ss.
 Synthetic.
 EP-571911-A.
 01-DEC-1993.
 24-MAY-1993; 108325.
 26-MAY-1992; US-889651.
 (BECT) BECTION DICKINSON CO.
 PI
 Shank DD, Spears PA;
 WI; 93-378844/48.
 New oligo/nucleotide probes specific for Mycobacteria - used for
 detection and amplification of Mycobacteria nucleic acid in
 samples
 Claim 3; Page 14; 23pp; English.
 Oligonucleotide probe MK14-A consists of nucleotides 5-95 of WK14
 (Q51735). It hybridized to all spp. of mycobacteria tested, but
 cross reacted to a few non-mycobacterial spp. The probe may
 be useful as an initial screen for mycobacterial infection.
 See also Q51735-45 and Q51747-59.
 Sequence 91 BP; 5 A; 17 G; 15 G; 4 T;

dbb 11 ssvhsyyvvhvshhsvhvhvhvsvvvvhvhvhvhvyhyvsy 60
::: ::::: : : : : : : : : : : : : : : :
718 GGGACCTGGAGGTTTGGTCAGCAGACACAGAAATCTTGCTGAATCCC 669

RESULT 6
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta galactosidase alpha-fragment.
OE E.coli beta galactosidase alpha-fragment; base substitutions; ss.
KW

Key	Location/Qualifiers
misc feature	19..69
/tag= a	
/function=	multiple cloning site
primer bind	187..204
/tag= b	
EP-285123-A	
05-MAY-1988	
30-MAR-1988	105163.
03-APR-1987	US-034819.
(SUUS) SUOMEN SOKERI OY.	
Lehtovaara P, Knowles J, Koivula A,	
WPI: 88-279927/40.	

Introducing random point mutations into nucleic acids - by prep of single stranded template, annealing a primer, misincorporation, completion of molecules and screening. Disclosure, Pt English.

Random point mutations were introduced into the alpha fragment of E.coli beta-galactosidase. The wild type sequence was obtained as a single stranded template and an oligonucleotide was hybridised to it to generate a popn of DNA molecules which terminate at all possible nucleotide positions within a specified region. The variable 3' ends generated in this way are used as primers for reverse transcriptase. Nucleotides are misincorporated by the transcriptase and the molecules are completed to forms that can be amplified and then expressed in a suitable host-vector system. The sequence covers all 176 different base substitutions, most of which occurred singularly in any given mutant.

```

CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

DB 1; Score 40; Match 10.2%; QryMatch 3.8%; Pred. No. 4.88e-09;
Matches 10; Conservative 52; Mismatches 36; Indels 0; Gaps 0;

Db 93 hyrrmrnbvrdynrdaawcyccrysvkycynachdhdyvbbvynvnhnnn 152
::: ::::: | | : : ::::: | | : : ::::: | | : : ::::: | | : :
Cp 289 CCATCACTCCCTCCAGACAGAGGCTCACTCTCCAGAAATGCTGTGCTTGGTCTCCT 230
|| : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 153 cecbnnhvcnbnhrnwnayvhrdarddhvcevh 190
|| : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 229 CCATCTGGGTTTTCATTCTCCCAAGTAAAGTCGACA 192

RESULT 7
ID Q35072 standard; DNA; 565 BP.
AC Q35072;
DE HCV envelope region nucleic acid.
KW Envelope; region; type C; hepatitis; virus; HCV; vaccine; serum;
KW non-A, non-B; amplify; ss.
OS Hepatitis C virus.
PN J04349885-A.
PD 04-DEC-1992.
PR 29-MAY-1991; JP-152169.
PA (TEIJ) TEIJIN LTD.
DR WPI; 93-022708/03.
PT Envelope region nucleic acid fragment - for type C hepatitis
virus (1), for producing vaccine
PS Claim 1; Page 2; 13pp; Japanese.
CC This sequence encodes a novel envelope region of type C hepatitis
virus (HCV). This fragment can be used for the preparation of a
CC vaccine for hepatitis C. This fragment was prepared from the serum
CC of non-A, non-B hepatitis patients and the envelope region DNA was
CC amplified by PCR using the primer sequences given in Q35073-76.
SQ Sequence 565 BP; 61 A; 92 C; 106 G; 85 T;

DB 6; Score 29; Match 25.3%; QryMatch 2.7%; Pred. No. 2.60e-03;
Matches 22; Conservative 36; Mismatches 29; Indels 0; Gaps 0;

Db 296 cyrcyarraybyavvyccrbcrygmrmwcbqbcayrtcgaytgcyythg 355
|::|::: ::::: |::|::: |::|::: |::|::: |::|::: |::|:::
Qy 271 CTGCTGAGGAGTATGTCAGACAGGGGCAACTGGGACCCACTTGCCTCATCCCTC 330
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 356 ggravgcyrcyytctgycsgcyvtst 382
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 331 CTGGGGCAGCTTCTGACAGGTCGCT 357

RESULT 8
ID N50030 standard; DNA; 501 BP.
AC N50030;
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN47.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..501
FT EP-163993-A.

```

```

PD 11-DEC-1985.
PF 17-MAY-1985; 105750.
PR 17-MAY-1984; GB-012564.
PA (SEAR) SEARLE G D & CO.
PI Bell ID, Boseley PC, Porter AG;
DR WPI; 85-311944/50.
DR P-PSDB; P50029.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2h, page 39; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mlxts.
SQ Sequence 501 BP; 110 A; 31 C; 69 G; 79 T;

DB 3; Score 28; Match 35.5%; QryMatch 2.6%; Pred. No. 7.98e-03;
Matches 33; Conservative 29; Mismatches 30; Indels 1; Gaps 1;

Db 45 ycartgcaraarytbtgcgcyrtbaaygngmdyrbqartaytgytbaargaym- 103
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp 845 CCTGCTGTGATGTTCTCTCAGAAATGTCGGGGCTCCTAGGGTCTCGTGGGTCCA 786
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db 104 gdcaygaytymgdathcncargarttyg 136
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp 785 GGAAAGAGTCCAGCAGCTCCATTCAAGAGTTTCG 753

RESULT 9
ID Q51731 standard; DNA; 10596 BP.
AC Q51731;
DE 31-MAY-1994 (first entry)
DE Plasmid pCisEBON for subcloning huHGF variants.
KW Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT enhancer 1..611
FT /*tag= a
FT /note= "CMV enhancer/promoter"
FT promoter 758..775
FT /*tag= b
FT /label= SP6 promoter
FT /misc feature 845..849
FT /*tag= c
FT /note= "SP6 RNA start"
FT /misc feature 902..966
FT /*tag= d
FT /function= cloning_linker
FT /polyA_signal 967..1107
FT /*tag= e
FT /note= "SV40 poly A"
FT /misc feature 1108..1531
FT /*tag= f
FT /function= SV40_origin
FT /misc feature 1580..4189
FT /*tag= g
FT /label= EBNA-1
FT /misc feature 4190..6374
FT /*tag= h

```


[illegible]

DT	04-SEP-1991 (first entry)	
DE	Sequence encoding new modified human beta interferon polypeptides	
DE	INFX 448	
KW	Antiviral; cell growth regulator; immune system regulator;	
KW	antiproliferative; ss.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	1..501
FT	/*tag= a	
PN	EP-163993-A.	
PD	11-DEC-1985.	
PF	17-MAY-1985; 105750.	
PR	17-MAY-1984; GB-012564.	
PA	(SEAR) SEARLE G D & CO.	
PI	Bell LD, Boseley PC, Porter AG;	
DR	WPI; 85-311944/50.	
DR	P-PSDB; P50030.	
PT	New modified human beta interferon polypeptide(s) - prepd. by	
PT	plasmid transformed bacteria, with improved antiviral,	
PT	anti-proliferative and immune regulating actions	
PS	Claim 28; Chart 2i, page 40; 71pp; English.	
CC	Compared with interferon beta prepd. by recombinant methods, the	
CC	INFs of the invention are more active and have different affinities	
CC	for cell surface receptors (allowing selective targeting); they	
CC	have higher therapeutic index; improved stability against microbial	
CC	breakdown during synthesis; and better in vivo solubility and	
CC	stability. They are also easier to recover from incubation mixts.	
SQ	Sequence 501 BP; 110 A; 30 C; 69 G; 80 T;	
DB	3; Score 25; Match 32.8%; QryMatch 2.4%; Pred. No. 2.09e-01;	
Matches	21; Conservative 23; Mismatches 20; Indels 0; Gaps 0;	
Db	45 ycartgcaraarytbybtggcarytbaaygmngdytcbgartaytgytbaargaymg 104	
Cp	CCTGTCGTGATGTTCTTCAGGAAATGTCGGGGCTCTTAGGGTCTCGGTGAGGTCCA 786	
Db	105 daar 108	
Cp	: :	
Cp	785 GGAA 782	
RESULT	13	
ID	Q58731 standard; cDNA; 1956 BP.	
AC	Q58731;	
DE	29-SEP-1994 (first entry)	
DE	Encodes secretory alkaline phosphatase reporter protein.	
KW	Multicistronic expression unit; bicistronic vector system;	
KW	recombinant protein production; reporter protein; SEAP;	
KW	human placental alkaline phosphatase; ss.	
OS	Homo sapiens (placenta).	
FH	Key	Location/Qualifiers
FT	CDS	43..1563
FT	/*tag= a	
FT	/product= human SEAP	
FT	mat peptide	_94..1560
FT	/*tag= b	
PN	W09405785-A.	
PD	17-MAR-1994.	
PF	26-AUG-1993; E02294.	
PR	27-AUG-1992; DE-228458.	
PA	(BEIE) BEIERSDORF AG.	
PA	(GBFB) GBF GES BIOTECH FORSCHUNG GMBH.	
PI	Achterberg V, Dirks W, Dorschner A, Eichner W, Hauser H;	
PI	Meyer-Ingold W, Mielke H, Wirth M, Dorschner A;	

DR WPI; 94-101190/12.
PT p-PSDB; R50010.
PR New multicistronic expression units - for producing equimolar
PT ants. of polypeptide(s) in mammalian cells as hosts
PS Example 1; Page 59-61; 109pp; German.
CC Reporter genes coding for secretory alkaline phosphatase (SEAP)
CC and for luciferase can be co-expressed using a bicistronic system.
CC The SEAP is secreted due to the introduction of a stop codon at
CC position 489 of the human placental AP sequence.
SQ Sequence 1956 BP; 380 A; 659 C; 584 G; 333 T;

DB 10; Score 25; Match 79.1%; QryMatch 2.4%; Pred. No. 2.09e-01;
Matches 34; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 1620 gccactgtccctgagtgctcccgcctcctggcgctcctgettc 1662
|| ||||||| || ||||| || ||||| ||
Cp 267 GTCTACTGCTCCAGATGCTCTGGCTTGTCTCTCTCATC 225

RESULT 14
ID Q35072 standard; DNA; 565 BP.
AC AC Q35072;
DT 20-MAY-1993 (first entry)
DE HCV envelope region nucleic acid.
KW Envelope; region; type C; hepatitis; virus; HCV; vaccine; serum;
KW non-A, non-B; amplifi; ss.
OS Hepatitis C virus.
PN J04349885-A.
PD 04-DEC-1992.
PE 29-MAY-1991; 152169.
PR 29-MAY-1991; JP-152169.
PA (TEIJ) TEIJIN LTD.
DR WPI; 93-022708/03.
PT Envelope region nucleic acid fragment - for type C hepatitis
PT virus (I), for producing vaccine
PS Claim 1; Page 2; 13pp; Japanese.
CC This sequence encodes a novel envelope region of type C hepatitis
CC virus (HCV). This fragment can be used for the preparation of a
CC vaccine for hepatitis C. This fragment was prepared from the serum
CC of non-A, non-B hepatitis patients and the envelope region DNA was
CC amplified by PCR using the primer sequences given in Q35073-76.
SQ Sequence 565 BP; 61 A; 92 C; 106 G; 85 T;

DB 6; Score 24; Match 28.8%; QryMatch 2.3%; Pred. No. 5.98e-01;
Matches 15; Conservative 23; Mismatches 14; Indels 0; Gaps 0;

Db 435 ctmbcmymgsmdyaygdgarrdyrcargytgyaaygtcdmtyatsecy 486
|| ::||::|| ::||:: : ::||:: : ::||:
Cp 339 CTGCCCGAGGAGGATGAGAGGCAGTGCGTCACAGTTGTCCCGCTGTGCC 288

RESULT 15
ID N50029 standard; DNA; 501 BP.
AC AC N50029;
DT 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE INFX 446.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..501
FT /tag= a
PN EP-163993-A.

```

PD 11-DEC-1985.
PF 17-MAY-1985; 105750.
PR 17-MAY-1984; GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Roseley FG, Porter AG.
PT WPI; 85-311944/50.
DR p-PSDB; P50028.
PT New modified human beta interferon polypeptide(s) - prep'd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PT Claim 28; Chart 2g, page 38; 71pp; English.
CC Compared with interferon beta prep'd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 112 A; 31 C; 69 G; 79 T;

DB 3; Score 24; Match 33.9%; QryMatch 2.3%; Pred. No. 5.98e-01;
Matches 20; Conservative 21; Mismatches 18; Indels 0; Gaps 0;

Db 45 ycartgycaraarybtbtgcgcyarbaayggngmddybtgartaitygytbaargaym 103
: : ||| : | : ||| : | : || : | : ||| : | : ||| : | : ||| : | :
Cp 845 CCGTGTCTGATGTTCTTGAGGAAATGCCGGGCGCTCCTAGGGTCCTGGGTGAGGGTCC 787

Job completed: Thu Sep 28 00:18:22 1995
Job time : 73 secs.

```

WAPSELEH (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nm n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Sep 28 00:06:22 1995; MasPar time 567.91 Seconds
1091.492 Million cell updates/sec

Tabular output not generated.

Title: >US-08-252-491-18
Description: (1:1062) from US08252491.seq
Perfect Score: 1062
N.A. Sequence: 1 ATCGAGCTGACTGAATTGCT.....ATCTCTCAGGAGGGTAA 1062
Comp: TACCTCGACTGACTTAAGCA.....TAGACAGAGCTCTTCCCAT

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 267821 seqs, 291840210 bases x 2

Database: emb1-new6

- 1 BCT
- 2 EST1
- 3 EST2
- 4 EST3
- 5 EST4
- 6 EST5
- 7 FUN
- 8 INV1
- 9 INV2
- 10 MAM
- 11 ORG
- 12 PIN
- 13 PRI
- 14 PRO
- 15 ROD
- 16 STS
- 17 SYN
- 18 UNC
- 19 VRT
- 20 VIR

Database:

- genbank89
- 21 BCT1
 - 22 BCT2
 - 23 BCT3
 - 24 BCT4
 - 25 BCT5

- 26 BCT6
- 27 INV1
- 28 INV2
- 29 INV3
- 30 INV4
- 31 INV5
- 32 MAM1
- 33 MAM2
- 34 PAT1
- 35 PAT2
- 36 PHG
- 37 PIN1
- 38 PIN2
- 39 PIN3
- 40 PIN4
- 41 PIN5
- 42 PIN6
- 43 PIN7
- 44 PRI1
- 45 PRI2
- 46 PRI3
- 47 PRI4
- 48 PRI5
- 49 PRI6
- 50 PRI7
- 51 PRI8
- 52 PRI9
- 53 ROD1
- 54 ROD2
- 55 ROD3
- 56 ROD4
- 57 ROD5
- 58 ROD6
- 59 ROD7
- 60 STR
- 61 STS1
- 62 STS2
- 63 STS3
- 64 STS4
- 65 SYN
- 66 UNA
- 67 VRL1
- 68 VRL2
- 69 VRL3
- 70 VRL4
- 71 VRL5
- 72 VRL6
- 73 VRT1
- 74 VRT2
- 75 VRT3

genbank-new6

- 76 BCT
- 77 EST1
- 78 EST2
- 79 EST3
- 80 EST4
- 81 EST5
- 82 EST6
- 83 EST7
- 84 EST8
- 85 INV
- 86 MAM
- 87 PHG

Qy 301 CAACCTGGGACCCACTTGCCTCTCATCCCTCCTGGGAGAGCTTCTTGACAGGTCGCTCTC 360

Db 576 ctccctggggccctgcagagccctcttggaaccagctctcccaaggcaggaccaca 635

Qy 361 CTCCTTGGGGCCCTGCAGAGCCCTCTTGGACCCAGCTTCTCCACAGGCGAGGACACA 420

Db 636 gctcacaaagatcccaatgccatctctgagcttccaaacactgctccagagaaagtg 695

Qy 421 GCTCACAGGATCCCAATGCCATCTCTCTGAGCTTCCACACAGCTCTCCGAGCAAGGTG 480

Db 696 cgtttcctgatgctgttaggaggtccacccctctgctcaggcgggcccacccaccaca 755

Qy 481 CGTTTCTGTATCTTCTAGAGGGTCCACCTCTGCTCAGGGGGGCCCCACCCACACA 540

Db 756 gctgtccccagagaacctctctagctctcaactgaacgagctcccaaacaggactct 815

Qy 541 GCTGTCCCGACGACGCTCTCTAGTCTCACACTGACGAGCTCCCAACAGGACTTCT 600

Db 816 ggattgtggagacaacttcactgctcagcagaactactggctctggctctctgaag 875

Qy 601 GCAITGTGGACAAACTTCACTGCTCAGCCAGACTACTGGGCTCTGGGCTTGAAG 660

Db 876 tggcagcaggattcagagccagattcctgctgctgaacaaacctccagggtccctg 935

Qy 661 TGGCAGCAGGATTCAGACCAAGATTCTGTGCTGTGCTGAACAACTCCAGGTCCTG 720

Db 936 gaccaaatccccgataacctgaacagatcacgaactcttgaatggaactcgtgactc 995

Qy 721 GACCAATCCCGGATACCTGAACAGATACAGAACTCTTGAATGGAACCTCGTGACTC 780

Db 996 tttcctggacctcacgcaggacctaggagcccgagacatttctcaggaacatcacag 1055

Qy 781 TTTCTGTGACCTCAGCGAGACCTTAGAGCCCGGACATTTCTCAGGACATCAGAC 840

Db 1055 acaggctccctggccaaccaacctccagctggatattctctcccaaccatcctcct 1115

Qy 841 ACAGGCTCCTGCGACCCAGCTTCCAGCTTGGATATTTCTTCTCCCAACCCATCCTCT 900

Db 1116 actgagcagatagctcttctctccctccaccaccttgccacacctgtgtccagctc 1175

Qy 901 ACTGGACGATATACGCTCTCTCTCCACCCACCTTGCCACCCCTGTGGTCAGCTC 960

Db 1176 caacctcgtctcctgacctctctgctcaacgcccacctaccagccctcttctaaac 1235

Qy 961 CACCCCTCTCTCTGACCTTCTGCTCCAGCCGCCACCTACCGACCTCTCTTAAC 1020

Db 1236 acatctacaccactcccagaattctgtctcaggaaggtaa 1277

Qy 1021 ACATCCTACACCCACTCCAGAACTCTCTCAGGAGGGTAA 1062

RESULT 3

LOCUS HSU11025 1341 bp mRNA PRI 15-OCT-1994

DEFINITION Human megakaryocyte growth and development factor (MGDF) mRNA, complete cds.

ACCESSION U11025

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Eukaryote crown group; Metazoa/Eumycota group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonataca; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1341)

AUTHORS Bartley,T.D., Bogenberger,J., Hunt,P., Li,Y.S., Lu,H.S., Martin,F., Chang,M.S., Samal,B., Nichol,J.L., Swift,S. et.al.

TITLE Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor Mpl

JOURNAL Cell 77 (7), 1117-1124 (1994)

MEDLINE 94291201

REFERENCE 2 (bases 1 to 1341)

AUTHORS Samal,B.B.

TITLE Direct Submission

JOURNAL Submitted (18-JUN-1994) Babru B. Samal, Developmental Biology, Amgen Inc., Amgen Center, Thousand Oaks, CA 91320, USA

COMMENT NCBI gi: 511223

FEATURES

source

1..1341

/clone="HFL-IVC7"

/organism="Homo sapiens"

/tissue_type="liver"

/dev_stage="fetal"

36..1097

/gene="MGDF"

/note="NCBI gi: 558078"

/codon_start=1

/product="megakaryocyte growth and development factor"

/translation="MELTELLIVWMLLTARLTLSPPACDLRLSKILRDSHLVH SRLSCPEVHPPTVPLPVPDVSIGEMKQMEETKAQDILGAVTLLLEGVMAARQL GPTCLSSLLGSGQVRLILGALQLGTLQPPQRTTAHDPNAIFLSPFHLLRGV RLFLVGGSTLCVRRPPTTAVPSRTSLVLTNLNPNRTSGLLFTNFSARTTSGSL LKXQGFRAKIPGLINQTSLSIDQPGYINRIHELINGTRGLFPGPSRRTLGAPDSS GTSDTGSLPNLPQGYSPSPHTPGQYTLPLPTLPVQVLPVQLHPLLPDSAPTPT TSPILNTSYTHSNLSQEG"

99..1094

/gene="MGDF"

/product="megakaryocyte growth and development factor"

mat_peptide

BASE COUNT 298 a 452 c 294 g 297 t

ORIGIN

DB 47; Score 1058; Match 99.8%; QryMatch 99.6%; Pred. No. 0.00e+00; Matches 1060; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 36 atggagctgactgaattgctcctcgtggtcatgcttctcctaactgcaaggctaacgctg 95

Qy 1 ATGAGCTGACTGAAATGCTCTGCTGGTCAATGCTTCTCTTAAGTCAAGGCTAACGCTG 60

Db 96 tccagcccgctcctcctcgtgtgactccgagctcctcagtaaaactgcttctgactcc 155

Qy 61 TCCAGCCCGGCTCCTCTGCTGTGACCTCCGAGCTCCAGTAAACTGCTTCGTGACTCC 120

Db 156 catgtccttccacagagactgagccagtgcccagaggttccacctttgctcaacctgtc 215

Qy 121 CATGCTTTCACAGACAGCTGAGCCAGTGCACAGAGGTTTCACTTTCCTTACACCTGTC 180

Db 216 ctgctcctcgtgtggaacttagctgggagaaatggaaacccagatggagagaccaag 275

Qy 181 CTGCTGCTGCTGTGGACTTTAGCTTGGGAGATGGAAACCCAGATGGAGGAGCAACAG 240

Db 276 gcacaggacattctggagcagtgacctctctgctggaggagtgatggcagcacggga 335

Qy 241 CCACAGCAGATTCTGGAGCAGTGAACCTTCTGCTGGAGGAGTGATGCCAGCAGGGGA 300

Db 336 caactgggaccacttgtctctcatcctcctcctggggagccttcttgacagcgtccgtc 395

Qy 301 CAACTGGGACCCACTTGGCTCTCATCCCTCTGGGGCAGCTTCTTTCGACAGGTCCTGTC 360

Db 396 ctctctggggccctgcagagcctctcttggaccagctcttccacagggcaggaccaca 455
|||||
Qy 361 CTCCTTGGGGCCCTGCAGAGCCTCTTGGAAACCCAGCTTCTCCACAGGCGCAGCACCA 420
|||||
Db 456 gctcacaaagatcccaatgccatctctctgagcttccaaacacctgctccgaggaaggtg 515
|||||
Qy 421 GCTCACAGGATCCCAATGCCATCTTCTGAGCTTCCACACCTGCTCCGAGGAAGGTC 480
|||||
Db 516 cgttctctgatgtgtgagaggttccacacctctgcgtcaggcgggccccaccaccaca 575
|||||
Qy 481 CTTTCTCTGATGCTTGTAGAGGGTCCACCTCTGCGTCAAGGGGGCCCCACCCACCA 540
|||||
Db 576 gctgtccccagcaaacctctctagtctccacctgaagagacctcccaacagagactct 635
|||||
Qy 541 CTTGTCGCCAGACAGACTCTAGTCTCTCACAGTGAACGAGCTCCCAACAGAGACTTCT 600
|||||
Db 636 ggattgtggagacaacttcaactgctcagccagagaaactactggtctgggtcttgaag 695
|||||
Qy 601 GGATTGTTGGACAAACTTCACTGCTCAGCCAGCAACTACTGCTGTGGGCTTCTGAAG 660
|||||
Db 696 tggcagcaggattcagagccaaagtctctgtgtgctgaacaaaacctccaggtccctg 755
|||||
Qy 661 TGGCAGCAGGATTACAGACCAAGATTCTGCTGTGCTGAACAAACCTCCAGGTCCTG 720
|||||
Db 756 gaccaaatccccgataacctgaacagataacagaaactcttgaatgaaactgtggactc 815
|||||
Qy 721 GACCAATCCCGGATACCTTGAACAGGATACACGAACCTTGAATGGAACCTGTGGACTC 780
|||||
Db 816 ttctctggacctcagcaggaccctaggagcccggaacatttctcaggaacatcagac 875
|||||
Qy 781 TTTCTGTGACCTTCAGCAGGACCTTAGGAGCCCGGACATTCTCAGGAACATCAGAC 840
|||||
Db 876 acaggtccctgccaccacccctcagcctggatattctcttccccaaccatctctct 935
|||||
Qy 841 ACAGGCTCGCTGCCACCAACCTCCAGCCTGGATATTTCTCTCCCAACCCATCTCTCT 900
|||||
Db 936 actggaacagatacagcttctctctctccacacaccttgccaccctgtggtccagctc 995
|||||
Qy 901 ACTGGACATATAGCTCTTCTCTTCCACCCACCTTGCCACCCCTGTGCTGCTCAGCTC 960
|||||
Db 996 caccctctgtctctgacctctctgtccaaagccaccacctaccagcctcttctaaac 1055
|||||
Qy 961 CACCCCTCTGCTTCTGACCTTCTGCTCCAAAGCCACCCCTTACGAGCCCTTCTTAAAC 1020
|||||

RESULT 4
LOCUS HUMTHROMA 6163 bp DNA PRI 29-JAN-1995
DEFINITION Human thrombopoietin gene, complete cds.
ACCESSION L36051
KEYWORDS thrombopoietin.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Homidae.
REFERENCE 1 (bases 1 to 6163)
AUTHORS Foster,D.C., Sprecher,C.A., Grant,F.J., Kramer,J.M., Kuijper,J.L.,
Holly,R.D., Whitmore,T.E., Heipel,M.D., Bell,L.A.N., Ching,A.F.,
McGrane,V., Hart,C., O'Hara,P.J. and Lok,S.
TITLE Human thrombopoietin: gene structure, cDNA sequence, expression,
and chromosomal localization

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (26), 13023-13027 (1994)
MEDLINE 95108091
COMMENT NCBI gi: 533214
FEATURES
source Location/Qualifiers
1..6163
/organism="Homo sapiens"
/sequenced_mol="DNA"
<1..158
/number=1
159..1826
/number=1
1827..1984
/number=2
join(1972..1984,2216..2343,2630..2716,4649..4816,
5053..5718)
/note="NCBI gi: 533215"
/codon_start=1
/product="thrombopoietin"
/translation="MELTELLIIVMLLTARLFLSSPAPPACDLRVLSKILRDSHYLH
SRLSQCEVRLPTPVLLPVDVFSIGENKQMEETKAQDILGAVTILLEGVMAARQQL
GPTCLSSLLGQSGVRRLLGALQSLGTLQPPQRTTAHKDENAIFLSFHLRGKV
RFIMLVGGSTLCVRRAPPTTAVPSRTSVLTINELNPTSGLTNTFTASARTTGSGL
LKWQGFRAKIPGLINQTSRLDQIPGYLARHELLINGTRGLFPGPSRTILGADLSS
GSTDGSLPPNIQPGYSPSTHPPTGQYTLFPLPTPLTPVQVLRHLLPDPSPATPTP
TSPILLNTSYTHSQNLSEQG"
1985..2215
/number=2
2216..2343
/number=3
2344..2629
/number=3
2630..2716
/number=4
2717..4648
/number=4
4649..4816
/number=5
4817..5052
/number=5
5053..56163
/number=6
BASE COUNT 1524 a 1706 c 1435 g 1498 t
ORIGIN

DB 51; Score 671; Match 99.0%; QryMatch 63.2%; Pred. No. 0.00e+00;
Matches 678; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 5034 gaccatctctgacctcagctctccacagggcaggaccacagctcacaaggtcccaa 5093
|||||
Qy 378 GAGCCTCTTGGAAACCCAGCTTCTCCACAGGCGAGCCACAGCTCACAGGATCCCAA 437
|||||
Db 5094 tgccatctctctgagcttccaaacacctgctccgaggaaggtgcggttctctgatgtgt 5153
|||||
Qy 438 TCCCATCTTCTGAGCTTCCAAACACCTGCTCCGAGGAAAGGTGCTTCTGATCTTCT 497
|||||
Db 5154 aggaggtccacctctgcgtcaggcggggccccaccaccacagctgtccccagcagaac 5213
|||||
Qy 498 AGGAGGTTCCACCTCTGCTGCTAGGGGGGGCCCCACCCACACAGCTGTCCCGCAGAAC 557
|||||
Db 5214 ctctctagtctcacactgaacagctcccaacacagagctcttgattgtggagacaaa 5273
|||||
Qy 558 CTCTCTAGTCTCTACACTGAACAGAGCTCCCAACAGGACTTCTGGATTCTTGGACACAA 617
|||||
Db 5274 ctctactgcctcagccagaactactggtctctgggcttctgaagtggcagcagggattcag 5333

Qy 618 CTTCACTGCGCTCAGCAGAACTACTGGCTCGGGCTTCTGAAGTGGCAGCGGATTTCAG 677
Db 5334 agccaagattcctgtctgtcgtgaacaaacctccaggtccctggaccataatccccggata 5393
Qy 678 AGCCAGATTCTGTCTGTCTGACCAAACTCCAGGTCCCTGGACCAATCCCGGATA 737
Db 5394 cctgaacaggatacacgaactcttgaatggaactcgtggactcttcttgaccctcaacg 5453
Qy 738 CTTGAACAGGATACCAACTCTTGATGGAACCTCGGACTCTTCTTGACCCCTCAGG 797
Db 5454 caggacctaggagcccgagacattctcctcaggaaacatcagacaggtccctgccacc 5513
Qy 798 CAGACCCCTAGAGCCCGGAGATTCTCTCAGAAACATCAGACAGCGCTCCCTGCCACC 857
Db 5514 caactccagctggatattctctctcccaaccatcctcctactgacagtatacgt 5573
Qy 858 CAACCTCCAGCCTGGATATTCTCTCTCCCAACCATCTCTTACTTGACAGTATAGCCT 917
Db 5574 cttcctcttccaccaccccttgcacccctgtggtccagctccacccctgttctctga 5633
Qy 918 CTTCCCTCTTCCAGCCACCTTGGCCAGCCCTGTGTCAGTCCACCCCTGCTTCTGA 977
Db 5634 cctctgtctcaacgcccacccctaccagcctcttcttaaacacatcctacacccactc 5693
Qy 978 CCTTCTGTCTCAAGCCGACCCCTTACCAGCCCTCTTCTAAACATCTTACACCCACTC 1037
Db 5694 ccagaatctgtctcaggagggtaa 5718
Qy 1038 CCAGATCTGTCTCAGAGGGTAA 1062

RESULT 5 HUMTA 7666 bp DNA PRI 03-MAR-1995
LOCUS Human gene for thrombopoietin.
DEFINITION D32046
ACCESSION
KEYWORDS thrombopoietin.
SOURCE Homo sapiens blood DNA, clone lambdaHGT1.
ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
REFERENCE 1 (bases 1 to 7666)
AUTHORS Sohma,Y., Akahori,H., Seki,N., Hori,T., Ogami,K., Kato,T.,
Shimada,Y., Kawamura,K. and Miyazaki,H.
TITLE Molecular cloning and chromosomal localization of the human
thrombopoietin gene
JOURNAL FEBS Lett. 353 (1), 57-61 (1994)
MEDLINE 95010765
COMMENT Submitted (08-Jul-1994) to DDBJ by:
Yoshiaki Soma
Pharmaceutical Research Laboratory
Kirin Brewery Co., Ltd.
2-2 Souja-machi 1 chome
Maebashi
Gunma 371
Japan
Phone: 0272-54-8618
Fax: 0272-52-2307.
NCBI gi: 577319
Location/Qualifiers
source 1..7666
/organism="Homo sapiens"
/sequenced_mol="DNA"

exon /tissue type="blood"
1420..1486
/number=1
exon /label=exon1
3157..3314
/number=2
/label=exon2
join(3302..3314, 3546..3673, 3960..4046, 5979..6146,
6383..7048)
/note="NCBI gi: 577320"
/codon_start=1
/product="thrombopoietin"
/translation="MELITELLIVMILLITARTLTLSPAPPACDLRVLSKILRSHVILH
SRUSQCEVHPDPTVPLIPAVDFSLGEMKQMEETRAQDIILGAVTLLEGVMAARQGL
GPTLSSILGQLSGQVRLILCALQLLQTPQRTTAHKDPNAIFLSFHLIRGKV
RFIMLVGGSTLCVRAPPTTAVSRSTSLVLTINELPRTSGLLTFTASARTTSGEL
LKWQGFRAKIPGLINQTSRLDQIPGYINRIHELINGTRGLFPGPSRRTLGADISS
GTSDTGSLPNLQPGYSPSTHPPTQYTLFPLPTLPVVQLHPLLDPDSAPTTP
TSPLLNTSYTHSQNLSEQ"
3546..3673
/number=3
exon /label=exon3
3960..4046
/number=4
exon /label=exon4
5979..6146
/number=5
exon /label=exon5
6383..7576
/number=6
/label=exon6
BASE COUNT 1897 a 2154 c 1807 g 1808 t
ORIGIN chromosome 3q27.

Db 51; Score 671; Match 99.0%; OryMatch 63.2%; Pred. No. 0.00e+00;
Matches 678; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 6364 gaccatcctctgcccctcagcttctccacaggcaggaccacagctcacaagatcccaa 6423
Qy 378 CAGCCTCTTGGAAACCCAGCTTCTCCACAGGCGAGCCACAGCTCACAAGGATCCCAA 437
Db 6424 tgcacatcttctgagcttccaaacacctgctccagagaaaggtgcgttctctgatgctgt 6483
Qy 438 TCCCATCTTCTGAGCTTCCAAACACCTGCTCCGAGGAAAGGTGCTTCTGATGCTGT 497
Db 6484 aggaggttccacctctcgctcaggcgggcccccacccacacagctgtccccacagaaac 6543
Qy 498 AGGAGGTCCACCTCTCGCTCAGGGGGGCCCCACCCACACAGCTGTCCCCACGACAC 557
Db 6544 ctctctagctcacactgaacgagctcccaacaggactctggattgttggagacaaa 6603
Qy 558 CTCTCTAGTCTCAGACTGACAGGCTCCCAACAGGACTTCTGGATTGTTGGAGACAAA 617
Db 6604 ctctactgctcagcagaaactactggctctgggcttctgaagtggcagcagggattcag 6663
Qy 618 CTTCACTGCTCAGCCAGACTACTGGCTCTGGGCTTCTGAAGTGGCAGCAGGATTTCAG 677
Db 6664 agccaagattcctggtctgctgaacaaacctccaggtccctggacaaaatccccggata 6723
Qy 678 AGCCAAGATTCTCTGCTGCTGAACCAACCTCCAGGTCCTCGACCAAAATCCCGGATA 737
Db 6724 cctgaacaggatacacgaactcttgaatggaactcgtggactcttcttgaccctcaacg 6783
Qy 738 CCTGAACAGGATACAGAACTCTTGAATGGAACCTGCTGGACTCTTTCTGGACCTCAGC 797

Db 6784 caggaccctaggagcccgagcatttctcctcaggaaacacagacagagctccctgcacc 6843
 Qy 798 CAGACCTTAGAGCCCGCGAGATTCTCTCAGAACATAGACAGGCTCCCTGCCACC 857
 Db 6844 caactccagctggattctcctctcccaacccatctcctactgagcagatatacgt 6903
 Qy 858 CAACTCCAGGCTGGATATCTCTTCCCAACCCATCTCTTACTTGGACAGTATAGCT 917
 Db 6904 ctctctctccaccacacttgcacccctgtgtccagctccacccctgtctctga 6963
 Qy 918 CTCTCTCTTCCACCCAGCTTGCACCCCTGTGTGTCAGCTCCACCCCTGTTCTGA 977
 Db 6964 cctctctgtccacagccacccctaccagccctctctaaacacatcttaacccactc 7023
 Qy 978 CCTTCTGTCTCCAGCCACCCCTTACAGCCCTCTCTTAACACATCTTACACCCATC 1037
 Db 7024 ccagaatctgtctcaggaaggtaa 7048
 Qy 1038 CCAGAATCTGTCTCAGGAGGTAA 1062

RESULT 6
 LOCUS MUSTHROA 1486 bp mRNA ROD 06-JUL-1994
 DEFINITION Mus musculus thrombopoietin mRNA, complete cds.
 ACCESSION L34169
 KEYWORDS thrombopoietin.
 SOURCE Mus musculus cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
 Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
 REFERENCE
 AUTHORS Lok S., Kaushansky, K., Holly, R.D., Kuijper, J.L., Lofton-Day C.E.,
 Oort, P.J., Grant, F.J., Heipel, M.D., Burkhead, S.K., Kramer, J.M.,
 Bell, L.A.N., Sprecher, C.A., Blumberg, H., Johnson, R., Prunkard, D.,
 Ching, A.F.T., Mathewes, S.L., Bailey, M.C., Forstrom, J.W.,
 Buddle, M.M., Osborne, S.G., Evans, S.J., Sheppard, P.O.,
 Presnell, S.R., O'Hara, P.J., Hagen, F.S., Roth, G.J. and Foster, D.C.
 Cloning and expression of murine thrombopoietin cDNA and
 stimulation of platelet production in vivo
 JOURNAL Nature 369, 565-568 (1994)
 MEDLINE 94261207
 COMMENT NCBI gi: 508540
 FEATURES
 source Location/Qualifiers
 1..1486
 /organism="Mus musculus"
 /sequenced_mol="cDNA to mRNA"
 174..1244
 /note="NCBI gi: 508541"
 /codon_start=1
 /function="c-mpl receptor ligand"
 /evidence=experimental
 /product="thrombopoietin"
 /translation="MELIDLLIAMLLAVRLTLSSVPAPCDPRLINKLLRDSLHL
 SRLSQCPDVP LSLIPVLPVDFSLGEMKTOTQSKAQDILGAVSLILLEGVMAARGOL
 EPGSLSSLLQGLSQVRLLLGALQGLLGTQPLQCRTHAKDPNALFLSLQQLRGKV
 RFLLVGEPFLICVRLPTLTPAVPSSTQLLTINKFPNRTSGLLIETNE SVTARTAGPL
 LSRLLQCFWKITPGLQNLNTRSPVQISGYLARTHPVNGTHGIFAGTSLQTLASDIS
 PCAFNKGSLAFNLQGLGPPSPSLAPDGHPTFPSPALPTTHGSPQPLHP LPDPSTTM
 PNSTAPHPVTHFPHRNLSQET"

BASE COUNT 331 a 484 c 337 g 334 t
 ORIGIN

DB 57; Score 602; Match 80.4%; QryMatch 56.7%; Pred. No. 0.00e+00;

Matches 856; Conservative 0; Mismatches 200; Indels 9; Gaps 4;

Db 174 atggagctgactgatttctcctggcgccatgcttcttgagtggaagactaactctcg 233
 Qy 1 ATGAGCTGACTGAATTGCTCTGCTGCTGCTGCTTCTCTTAACGCAAGGCTAACGCTG 60
 Db 234 tccagcccgtagctcctgctgtgaccccgagactcctaaataaactgctgctgactcc 293
 Qy 61 TCCAGCCCGGCTCTCTGCTGTGTGACTTCCGAGCTCCAGTAACTGCTTCTGCTACTCC 120
 Db 294 caactccttcacagcgactgagtcagtgctccgagctcgacacctttgtctatccctgtt 353
 Qy 121 CATGCTCTTCACAGCAGACTGAGCCAGTGGCAGAGGTTACCCCTTTGCTTACACCTGTC 180
 Db 354 ctgctgctgtgtggaactttagcctgggagaatggaaaccccgagcggaacagagcaag 413
 Qy 181 CTGCTGCCCTGTGTGACTTTAGCTTGGAGAGATGGAAAACCCAGATGGAGGAGCAAG 240
 Db 414 gcacgagcattctaggggcagtgctccttctactgagggagtgatggcagcagcagga 473
 Qy 241 GCACAGGACATTCTGGAGCAGTGACCTTCTCTGGAGGAGTGATGGCAGCACGGGA 300
 Db 474 cagtggaacccctcgcctctcatccctcctgggacagcttctgggcaggttcgcctc 533
 Qy 301 CAATGGGACCCACTTGCCTCTCATCCCTCTCTGGGCGAGCTTCTTGACAGAGCTCCGTC 360
 Db 534 ctcttggggccctgcagggctccttaggaacccagcttctctacagggcaggaacaca 593
 Qy 361 CTCTTGGGGCCCTGCAGAGCTCTTGGAAACCCAGCTTCTCTCCAGGGCAGGACCA 420
 Db 594 gctccagagggccccaatgccctctcttctgaggttgcaacaactgtctgggggaaggtg 653
 Qy 421 GCTCACAGGATCCCAATGCCATCTCTCTGAGCTTCCACACCTGTCTCGAGGAAAGGTG 480
 Db 654 cgctctctgcttctgtagaaggtccaccctctgtgcagagggacctgccaaacaca 713
 Qy 481 CTTTCTCTGATGCTGTAGGAGGTCACCCCTCTGGCTCAGGGGGCCCAACCAACCA 540
 Db 714 gctgtcccaacagtagtacttccactcctcacactaaacaggttcccaacagagactct 773
 Qy 541 GCTGTCCCAACAGACCTCTCTAGTCTCTCAGTGAACGAGCTCCCAACAGGACTTCT 600
 Db 774 ggattgtggagacgaacttcagtgacagccagaactgctggccctggactctcgagc 833
 Qy 601 GGAATTCTTGGAGACAACTTCACTGCTGCAGCCAGAACTACTGGCTCTGGGCTTCTGA 660
 Db 834 aggttcaggaggttcagagtgacttactcctggtcagctaaatcaaacctccaggtcc 893
 Qy 661 TGGCAGCAGGATTCAGAGCCCAAGATT---CCTGGTCTGCTGAACCAAACTCCAGGTC 717
 Db 894 ccagtcctcaaatctctggataacctgaacggacacgagcctgtgaatggaactcatggg 953
 Qy 718 CTGACCAAAATCCCGGATATCCTGAAACAGGATACAGCAACTCTTGAATGGAACTCTGGA 777
 Db 954 ctcttctgtggaactcacttcagacccctggaagcctcagacatctcgccggagcttcc 1013
 Qy 778 CTCTTCTTGGACCCCTCAGCAGGACCCCTAGGAGCCCGGACATTTCTCTCAGGAACATCA 837
 Db 1014 aacaaaggctccctggcattcaacctccagggttgacttctccttctccaaagccttget 1073
 Qy 838 GACAGAGGCTCCTGCGCCCAACCTCCAGGCTGGATATTCTCTTCCCAACCACTCCT 897
 Db 1074 cctgagtgacacacaccccttccctcttcaactgacctgccccaccacccatggatctcca 1133

DB 57; Score 602; Match 80.4%; QryMatch 56.7%; Pred. No. 0.00e+00;

AUTHORS Baer, R.J., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G.F., Hudson, G.S., Satchwell, S.C., Sequin, C., Tuffnell, P.S. and Barrell, B.G.

TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome

JOURNAL MEDLINE Nature 310 (5974), 207-211 (1984)

REFERENCE 84270667

AUTHORS 2 (bases 1 to 172281)

TITLE Sequence analysis and in vitro transcription of portions of the Epstein-Barr virus genome

JOURNAL MEDLINE J. Cell. Biochem. 19 (3), 267-274 (1982)

REFERENCE 83109311

AUTHORS 3 (bases 1 to 172281)

TITLE Latent and lytic cycle promoters of Epstein-Barr virus

JOURNAL MEDLINE EMBO J. 2, 1331-1338 (1983)

REFERENCE 5 (bases 142687 to 159853)

AUTHORS Bankier, A.T., Deininger, P.L., Farrell, P.J. and Barrell, B.G.

TITLE Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus

JOURNAL MEDLINE Mol. Biol. Med. 1 (1), 21-45 (1983)

REFERENCE 85035713

AUTHORS 6 (bases 112620 to 125316)

TITLE DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus

JOURNAL MEDLINE Mol. Biol. Med. 1 (3), 369-392 (1983)

REFERENCE 85060424

AUTHORS 7 (bases 159853 to 172281)

TITLE DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences

JOURNAL MEDLINE Mol. Biol. Med. 1 (4), 425-445 (1983)

REFERENCE 85060428

AUTHORS 8 (bases 87650 to 92703)

TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus

JOURNAL MEDLINE EMBO J. 3 (5), 1083-1090 (1984)

REFERENCE 84236104

AUTHORS 9 (bases 76089 to 79808)

TITLE Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes

JOURNAL MEDLINE Nucleic Acids Res. 12 (12), 5087-5099 (1984)

REFERENCE 84247360

AUTHORS 10 (bases 1 to 172281)

TITLE Unpublished

JOURNAL MEDLINE Hatfull, G.F., Barrell, B.G., Quinn, J. and McGeoch, D.

REFERENCE 11 (bases 1 to 172281)

AUTHORS Arrand, J.R., Rymo, L., Walsh, J.E., Bjorck, E., Lindahl, T. and Griffin, B.E.

TITLE Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments

JOURNAL MEDLINE Nucleic Acids Res. 9 (13), 2999-3014 (1981)

REFERENCE 82014887

AUTHORS 12 (bases 1 to 172281)

AUTHORS Kozak, M.

TITLE Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes

JOURNAL MEDLINE Nucleic Acids Res. 9 (20), 5233-5262 (1981)

REFERENCE 82059504

AUTHORS 13 (bases 7315 to 9312)

TITLE A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells

JOURNAL MEDLINE Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)

REFERENCE 84222045

AUTHORS 14 (bases 45415 to 52824)

TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus

JOURNAL MEDLINE EMBO J. 3 (4), 813-821 (1984)

REFERENCE 84207939

AUTHORS 15 (bases 45644 to 52450)

TITLE Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript

JOURNAL MEDLINE J. Virol. 48 (1), 135-148 (1983)

REFERENCE 83294686

AUTHORS 16 (bases 1 to 172281)

TITLE Direct Submission

JOURNAL MEDLINE Submitted (05-JUN-1984) to the EMBL/GenBank/DBJ databases

REFERENCE 17 (bases 1 to 172281)

AUTHORS Bodescot, M. and Perricaudet, M.

TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs

JOURNAL MEDLINE Nucleic Acids Res. 15 (14), 5887 (1987)

REFERENCE 87289053

AUTHORS 18 (bases 1 to 172281)

TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome

JOURNAL MEDLINE EMBO J. 7 (3), 769-774 (1988)

REFERENCE 88283646

AUTHORS 19 (bases 1 to 172281)

TITLE Direct Submission

JOURNAL MEDLINE Submitted (18-MAR-1988) to the EMBL/GenBank/DBJ databases. Farrell, P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG

COMMENT CDS

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BAPF3 is the third leftward frame starting in Bam HI fragment A. BORF1 is the first rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met

codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES of POLYA signals

This feature lists all occurrences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog ATTAAA is only listed when it is found in a position close to the end of a major reading frame.

SITES of DONOR and ACCEPT sequences

This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case.

Restriction enzyme SITES.

Only the positions of the sites Bam HI (BAW) are listed.

RPT

This feature is used to define repetitive sequences.

SITE DEL

This feature defines deletions in B95-8 with respect to other strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN

Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

ORGRPL

Denotes the region that encompasses an origin of replication (ori P). [13].

NUMBERING

The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

NCBI gi: 59074	
FEATURES	Location/Qualifiers
source	1..172281
	/organism="Epstein-Barr virus"
	/strain="B95-8"
mRNA	58..272
mRNA	360..458
misc_feature	/note="exon 2 terminal protein RNA"
	complement(535)
	/note="polyA signal: AATAAA"
mRNA	540..788
	/note="exon 3 terminal protein RNA"
	complement(535)
mRNA	871..951
	/note="exon 4 terminal protein RNA"
mRNA	1026..1196
	/note="exon 5 terminal protein RNA"
promoter	/note="exon 6 terminal protein RNA"
	complement(1192)
mRNA	/note="TATA: TATAAAT"
	1280..1495
promoter	/note="exon 7 terminal protein RNA"
	complement(1383)
mRNA	/note="TATA: CATAAAA"
	1574..1682
promoter	/note="exon 8 terminal protein RNA"
	1676
promoter	/note="TATA: TATAAAG"
	1691
	/note="TATA: TATTAAA BN-R1 late promoter before BNF1, gives 4.1kb late RNA. Probably encodes non glycosylated 140kd protein in membrane antigen. Also two latent RNAs spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson et al, 1985). The longer one encodes terminal protein."
CDS	1736..5692
	/note="BNF1 reading frame, 5 NKT/5; NCBI gi: 59075"
	/codon_start=1
	/translation="MEERGRETMPEVARYGPFIMVRLFGDCEANIQEERLYELLSD PRSALGLDPCPLIAENILLVALRGNTNDPPQQRERELALVGLLNGEQEHLGCT ESAIEASGNNYYAYFGDMWAPSTWAEIQQFLRLGATYVLRVEMGRQFGEVHRS RPSFRQAINHLVFDNALRYDSCQAAFGQFALLVAGTADTRDRLKLNWVF GGRAAGGRLADELKIYVALRDYTSGLHVLQPTETLDTWKVLSRDRTRHSLHGFIH AAGTIQANCPQLFMRQHPGLFFVNAIASSLQWYIQTATGPDARAAAARQQAFQT RAAECHARSGVPVYAGFYRTNATKGGELQPTMFGELGAIKHQALDTRDYDGH YLIMLGPFPWSGLTAPPCCPYAESWQAQAVOTALELFSALYPAPCISGVARPCPSA VIEHGLSVKPGGLLLFLSHLPDVKDGLGEMPARATGPGMQQFVSSYFINPACSNV FITVRQRGEKINGRTVILQALGRACDMAGCQHYLVGSTVPLGGINFVNDLASPVSTAE MDDSPFTTVEFPPIQEEGASSPVLVDDESMDISPSVELPWLISLCLTSILSHPTV GSKEHLVRHTRDVRSGRVAQQPGVGPLDLFLADYAFVAHSQWTRPGGAPLPYRTWD RMTEKLLVRSKPGGENKVSQVITVLGEGYKVIDLREGTRLMAEALLNAACAPIL DPEDVLLTLHLHDPRADNSAVMEAMTAASDYARGLGKLVTFGSASCPETGSSASNF MTWVASVAPGECFCLITPVLQKTGSLIIAVRCDCGKIQGSLFEQLFSDVATTPRA PEAUSLKNLFRVAQQLVKSGIVLSGHDISDGLVTCILVEMALAGQGVITIMPVADSY LPFEMFAEHPGLVFEVEERSVGEVQLTRSNMYPVAVRGVGEQGDQMFQVHGPEYV LQKSLILLCTWSFASEQYECICLRDRINRSMHVSVDYCYNEALVSPITGKNLSRPL VTEPDRCQVAVLCAPGTGHSLSLAAFTNAGCLCRVFFREVRDNTFLDKYVGLAIG GVHGARDSALAGRATVALINRFPALRDAILKFINRPDTFSVALGELGVQVLAGLVAG


```
STNPPAPGVENVQSRSLIAPNASCMEFSRNLNISIPATTSVMIRGICGCVLPCW
VQSGCIGLQFTNLGMPYVLIQNAHQIACHFHSNGTDAMRFAMNYPNRPTEQNTAGLCS
RDCGRHIALLCDPSLCTDFQWEHIPPAPGCHPTGCSPTWLMFQAAHLWSLRHGRESE"
/notes="polyA signal: AATAAA"
3955
/notes="BAM: Bam H1 Nhet/h"
3994
/notes="BAM: Bam H1 h/C"
5408..5856
/notes="exon 9 terminal protein RNA"
5841
/notes="polyA signal: AATAAA, end of 4.1kb late RNA and TP
latent RNA."
5863
/notes="alternative end to TP cDNAs"
6097
/notes="TATA: TATAAGA"
6629..6795
/notes="Pol III RNA EBER 1"
complement(6823)
/notes="TATA: CATAAAT"
6956..7128
/notes="Pol III RNA EBER 2"
7315..9312
/notes="origin of replication, ori P (Yates et al, 1984,
1985)"
7421..8042
/notes="21x30bp repeats, binding sites for EBNA-1 (site I,
Rawlins et al, 1985). Tandem repeat part of orip (Reisman
et al, 1985). Also functions as a cell type specific
enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
7738
/notes="TATA: TATAAAT"
7888
/notes="TATA: TATAAAT"
8573
/notes="TATA: CATAAAT"
complement(8680)
/notes="polyA signal: AATAAA"
complement(8755)
/notes="polyA signal: AATAAA"
8962
/notes="polyA signal: AATAAA"
9021..9133
/notes="HPN: dyad symmetry, site II for EBNA-1 binding
(Rawlins et al, 1985). Dyad symmetry part of orip (Reisman
et al, 1985)"
complement(9398)
/notes="TATA: TATAAAT"
9631
/notes="TATA: TATAAAT BC-R1 late promoter before BCRF1"
9675..10187
/notes="BCRF1 reading frame; NCBI gi: 59076"
/codon start=1
/translation="MERRLVTVLQCLVLLYLAEPCGGTDQCDNFPOMLRDLRDFSRV
KTFQFKQDVNLLIKESLLEDFKYLGCQALSEMIFYLEEVMPOAENQDPEAKDQV
NSIGENLKTLLRLRCHRLFPCENKSKAVEQIKNAFNKLQEKGIYKAMSEFDIFYN
IEAYWTIKAR"
10076
/notes="TATA: GATAAAA"
complement(10148)
/notes="polyA signal: AATAAA"
10173
```

```
/notes="polyA signal: AATAAA"
10257
/notes="polyA signal: AATAAA, end of 0.8kb late RNA from
BCR1 and end of 1.6 kb late RNA, start unknown"
complement(10277)
/notes="polyA signal: AATAAA"
complement(10975)
/notes="TATA: CATAAAT"
11305
/notes="TATA: TACAAA; BCR2 promoter for highly spliced
EBNA latent RNAs."
11336..11480
/notes="exon C1 of Bodescot et al (1986) RNAs"
11524
/notes="TATA: TATAATT"
complement(11587)
/notes="polyA signal: AATAAA"
complement(11606)
/notes="TATA: CATAAAT"
11626..11657
/notes="exon C2 of Bodescot et al (1986) RNAs"
11796
/notes="TATA: TATAAGT"
complement(11799)
/notes="TATA: TATAAAA"
12001..15072
/notes="3072 repeat 1"
12541..13689
/notes="BCRF2 3072 repeat, reading frame 1"
13215
/notes="BAM: BamH1 C/W"
14352
/notes="TATA: TATAAAG BWR1 one of the promoters for highly
spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al,
1986)"
14384..14410
/notes="exon W0 of EBNA/LP RNAs"
14554..14619
/notes="exon W1 (also W66) part of leader protein (LP)
gene. LP is also called EBNA-5 (Dillner et al, 1986) and
EBNA-4 (Rowe et al, 1987)."
14559..14619
/notes="exon W1' (also W61) of EBNA/LP RNAs forms
initiator met when fused to exon W0 or exon C2."
14701..14832
/notes="exon W2 (also W132) part of LP gene"
15073..18144
/notes="3072 repeat 2"
15613..16761
/notes="BWRF1 reading frame 2"
16287
/notes="BAM: BamH1 W/W"
17424
/notes="TATA: TATAAAG"
17626..17691
/notes="Exon W1"
17773..17904
/notes="Exon W2"
18145..21216
/notes="3072 repeat 3"
18685..19833
/notes="BWRF1 reading frame 3"
19359
/notes="BAM: BamH1 W/W"
```

```

promoter      20496
/notes="TATA: TATAAAG"
mRNA          20698..20763
/notes="Exon W1"
mRNA          20845..20976
/notes="Exon W2"
repeat_region 21217..24288
/notes="3072 repeat 4"
CDS           21757..22905
/notes="BWRFL reading frame 4"
misc_feature   22431
/notes="BAM: BamH1 W/W"
promoter      23568
/notes="TATA: TATAAAG"
mRNA          23771..23835
/notes="Exon W1"
mRNA          23917..24048
/notes="Exon W2"
repeat_region 24289..27360
/notes="3072 repeat 5"
CDS           24829..25977
/notes="BWRFL reading frame 5"
misc_feature   25503
/notes="BAM: BamH1 W/W"
promoter      26640
/notes="TATA: TATAAAG"
mRNA          26842..26907
/notes="Exon W1"
mRNA          26989..27120
/notes="Exon W2"
repeat_region 27361..30432
/notes="3072 repeat 6"
CDS           27901..29049
/notes="BWRFL reading frame 6"
misc_feature   28575
/notes="BAM: BamH1 W/W"
promoter      29712
/notes="TATA: TATAAAG"
mRNA          29914..29979
/notes="Exon W1"
mRNA          30061..30192
/notes="Exon W2"
repeat_region 30433..33504
/notes="3072 repeat 7"
CDS           30973..32121
/notes="BWRFL reading frame 7"
misc_feature   31647
/notes="BAM: BamH1 W/W"
promoter      32784
/notes="TATA: TATAAAG"
mRNA          32986..33051
/notes="Exon W1"
mRNA          33133..33264
/notes="Exon W2"
repeat_region 33505..36576
/notes="3072 repeat 8"
CDS           34045..35193
/notes="BWRFL reading frame 8"
misc_feature   34719
/notes="BAM: BamH1 W/W"
promoter      35856
/notes="TATA: TATAAAG"
mRNA          36058..36123
/notes="Exon W1"

```

```

mRNA          36205..36336
/notes="Exon W2"
repeat_region 36577..39648
/notes="3072 repeat 9"
CDS           37117..38265
/notes="BWRFL reading frame 9"
misc_feature   37791
/notes="BAM: BamH1 W/W"
promoter      38928
/notes="TATA: TATAAAG"
mRNA          39130..39195
/notes="Exon W1"
mRNA          39277..39408
/notes="Exon W2"
repeat_region 39649..42720
/notes="3072 repeat 10"
CDS           40189..41337
/notes="BWRFL reading frame 10"
misc_feature   40863
/notes="BAM: BamH1 W/W"
promoter      42000
/notes="TATA: TATAAAG"
mRNA          42202..42267
/notes="Exon W1"
mRNA          42349..42480
/notes="Exon W2"
repeat_region 42721..45792
/notes="3072 repeat 11"
CDS           43261..44409
/notes="BWRFL reading frame 11"
misc_feature   43935
/notes="BAM: BamH1 W/W"
promoter      45072
/notes="TATA: TATAAAG"
mRNA          45274..45339
/notes="Exon W1"
misc_feature   45415..52824
/notes="DEL: DAUDI deletion (Jones et al, 1984)"
mRNA          45421..45552
/notes="Exon W2"
misc_feature   45644..52450
/notes="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
repeat_region 45793..47643
/notes="3072 repeat 12"
CDS           46333..47481
/notes="BWRFL reading frame 12"
misc_feature   47007
/notes="BAM: BamH1 W/Y"
mRNA          47761..47793
/notes="Exon Y1 Bodescot et al, 1984"
promoter      47831
/notes="TATA: TATAAGT"
mRNA          47878..47999
/notes="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and
Strominger, 1985), last common exon"
misc_feature   complement(48023)
/notes="polyA signal: AATAAA"
mRNA          48386..48444
/notes="exon Bodescot et al, 1984"
CDS           48386..50032
/notes="Coding exon for EBNA-2 (Sample et al, 1986)"
CDS           48429..49964
/notes="BYRF1, encodes EBNA-2 (Dambaugh et al, 1984;
Dillner et al, 1984)"

```



```

APAPSLPQQPPSPSARASPLPQQOPLSATPAPPAAQLPSPATLLEPKHP
PAADRGTEISPPSPGQOQSGDDASGGLVYLSLDEEFFLSMSDEEASDILAS
DIPITTEDEDEDEVENSLSESGSAPTSPTITLDTARSQYTOTTFDIETPMDFVPLE
SNIRTAGHTYQOQAIVDPASREVPEDALSHIDYLLVTVLEOGLIRSDRSSLV
NILEEFKDSMGLQVPTLIDLEQLNLTSEINTIANMLSENKGRAGEFKHKLAKLEAC
LPSLATKQAVNDAGAKMLAEIPQIAESDDGKFDLEARRRLDULLSGSDQOAGEGGG
EPEDNSIYRGVDPVPLVLDDESCKLLSLAEAAPTAVARQOQVDEEDVRFALILTA
IYEGAPPAASVPVYVHNVAVRNSAALHVRCTADIRKVASAASDYLSYLESLPT
VMDFDLDLHLRLETCQIITASLPILNIRYTSIEMDYRELLYLCTALSDMSGIPMLERV
EEDPSIAPPEETVANKOKELETTRENEKRLTILDDIEAMLGLAGVASACAPIS
PASPSATPANHDPENATPLADTAALTIPIVEKTYANAGSIVGAKNPTTYIRURTIQ
QIVRSKYIMNLIKSTIYITDNYIASFEESIDHLYDLPLDPVEQDGDIDRLIDPMV
SEALHTECMGNLITLEPARIVALQNFATHTLAKETAVALNLLPCLLAVYDATTIGQAP
EDARLLSGQLQNSOTLLPKLKKRFLSYLQKLKNNNDQLQKQVQWRLNEGEFK
PATYEQLEAFIDTAPNKEIKROYKRLQIMETGRKEKELREQEDKEKREARRAN
EAWARLXALGARPEPAPTSPDOWNTLLASLLPNDTSAAMAAAVARNTDIDLSLQ
ILAMLLGITVRREKRLSLVDDGGAERMAEAPGWFDTIETGFLARLAMPATPA
ATAGGGGCGRAEAGAAFRATAADAIIRSAQAOTRQALQSDPKMSAVVNTIDLEAPY
AYERGLAGLKEKRAAEALATVSEYVRTLPENATDPPQANILPPTTIPQATAPP
RLASDALSMPKPKQLITRREDDLOAQDFFSELLITEAEAEVRALEEQVRESQTIM
AKAHEMAASTRRGFTALAEAVLSRSHDEAPDDELRLSLIPSPKAPVQAPLEAALHRA
AGNSKPYRKSILAAAKIRGICEAVRGLSEGALALAGAGAWINLAAADGETHELTR
LLEVEGMAQNSDMGEELRLALATLDPKRVAGKETVADWKRLSLRLALIQEAQES
QUAGTLQDVLQARGHTDPRQJLIVVEARGALGASQSYALIKDKLILRYASAKS
FLAFYETQAPTPTVTVKHTNNLPULLITISAPPTGCMGNQAPTRRAQFLAAAGPARYAGT
LWLETESPCDIPNPAVVSADTOERLNYIPVYHNFLEYVMPVLENPEAFSLITPAGPQ
AIGCFQDDQERRRTLASVASRLSAAADSMTDTPVDSNAGELLREYVSAPKALM
EDLADNPITVAMTLLAHASLIASRNHPYPAPATDREVILLLEQREMMALLVGTHPAYAA
AFLGAPSYAGLINSALARDGGLDLSLVLYRLVRSPASGRGCMPESTTRCSNDG
EDARRLTRHLAGCTPGIFPDQAEWENDTRALWHPFEFLGLVHNQSTARRACMLL
LARCPAPALQQLMHSIRPLEGCVAFQDYLRDFVKQAYTRGEEIPRAEGLEVPRETP
SSYGTGTGVRRLNLPYGTPTGPKRSGDTPVSVFEAAVAALFGRPLTLFVSSQY
LEWIKTLQVRRVAPLLVCDGHSEPPRSIVETISINFLODLDCYSESEPEMSIFARQ
AWMLRELTENARAQKCARPITVALIANRKNIIKCFYIRHNLPDVQFYNAAGASR
WPTDVLNPFYEHEDPPLFVGQYQLPNPNRNVOELFSGFPVRVGHVSGDGFQSDNT
PASSDRLOQLGGEGTDOCEKSTTAESEASGPPSPOSPLLEKVAPGRDMLSPTSSP
RDVTVTGLAAPTITLPGRLMARPYGAETRASESPPRSGSPSPRPWKUSLELILQP
APQOSTGPGWASQGP IVYTLSPHSTPASGQKHTIQIQLGLVSOQPSYPPSAPY
PKGOSTGGIAPTSAASLITFGLQDQTOASSODPPYGHSTIMOREKKQGGREEAEI
RPSATRLTAVGLRPPAPVAVAGAAASATPAFDGENSGPPIPQAPALGSLAAPAH
PTGALAPRPQKQRPQDAAALPPTIKAVGARVPKATCALAAGARPGRQPTAAP
PSAASPPRVSLPVRSQOQSPAIPLPMHSGSEPGARPEVRLSQYRHRAGPOTTVRKE
APPSAASQLPMKPKCKDSNYPYPSGARYPAPFQALSFQSQVASPAPSSDQTLIMNT
PSYVTQFLIREVTGTSGLDVPVPSGSPSSLSSTAAPAEQDLRLYSLTLSQASRV
LSRFVSQLRRKLERSTHRLIADLERLKFYL"
62249
/notes="BAM: BamH1 F/Q"
62430..62477
/notes="Site III for EBNA-1 binding (Rawlins et al, 1985)"
66121
/notes="BAM: BamH1 Q/U"
67477..67649
/notes="Exon in EBNA-1 RNA (Speck and Strominger,1985) and
cDNA clone T4 (Bodescot et al, 1986)"
69410
/notes="BAM: BamH1 U/P"
69684..69930
/notes="5 x 51bp repeats"
70387..70521
/notes="9 x 15bp repeat"
70750
/notes="TATA: CATATAA"
complement(71520..75239)
/notes="BOLF1 reading frame, 1 NXT/S analogous to VZV RF

```

misc_feature

misc_feature

misc_feature

mRNA

misc_feature

repeat_region

repeat_region

promoter

CDS

```

21; NCBI gi: 832979"
/codon_start=1
/translating="MASAMESDSSGGSGCADAQPPLAEDVGGIARVTRQLLLSGDDPA
ARLRALMPLELIGFGLDGLAQPVIVRDFINTLTILMSGHYPAVALNHAYYILRAASF
SRRSFGLHLEALDVLASSLPPTTASPAYDDPJDGSLRIAEATRALAAAYRIIEGS
GEVLAVAGCTATTAFAVEELVADTYLARMDAFPREGLSFYAFNAAKTTLGRMLVTVYAE
TRNYPAAACQOQPTMAADIKAMAVELVHSGGAGGEGSGGLFRHPESLSVVA
SLPTLARRARVEILGYPAASGGQTPVNAVPVLAFDAARLRLLEPSGALFYDYVEAL
LMDQTYGVPDSVIEAFAGMAAMEALAAVQEAASGRASFSPAALTEQAVTLLSAGL
NETVAGDYAMMLASVPRSRWRLEATALESISCFALHFFRLIPTASPTSRFAR
VARAAYLRAEACAVDRRARRTSGPSTAAAPATAVGVGAADPMDAVTPRLIRIVPP
PAAETVEAGYCSLRAGQTRADAPAVVHSPFPAFTEEQWPAALREVLSNLPKTR
AAAPTTOVAGYCSLRAGQTRADAPAVVHSPFPAFTEEQWPAALREVLSNLPKTR
SLAALRVLSFGSDLAIPSELTRARRPLELIYATVMEIYDGNPMPGSPQAVGLRP
RLLYDAETGRVQIPLATEEEEAHVAVKESVSSPHYSTDLQTLKSWEGIQDVC
RDAARWALATADTATLRRLLVPAALRESGTADHPIMAHITSEP LRPDLEELNERVEH
ALELCYSLTCALRSSVARERDYTFARLQFPALDAERAEAIVRRDARPPVFIIPAR
RLPQCGADTPPLSMDDILYLKLSICKALVDVLDHHPAEPETTP IKTYTPAMDLAPEQ
ITVTRSPSVLAFAFARTARVQTHLVPALTDSPSPGVQTPPPFRILIPAKKLAAILLG
NORNAKSRASRLDLSPPHGRMARVLDSPPSFSSDSDQDEGEADLRGVGGG
GCGAEEDRERPSIDTAAABAQKVETSCPRRSPTTTPSPSRASGGGSPDRGEAEAH
TPPYLSAANAASVRPRTRGATRRPRPTADE"
complement(72192)
/notes="TATA: TATATAA before BOLF1"
73468
/notes="BAM: BamH1 P/O"
75017
/notes="TATA: TATTAA BO-R1 late promoter before BOLF1,
gives 3.9kb late RNA"
75238..76332
/notes="BOLF1 late reading frame, 2 NXT/S homologous to VZV
RF20; NCBI gi: 832978"
/codon_start=1
/translating="MKVQSGSDRRRLQRRITAGLLPPPARLINISRGSEFTDRVGRIVE
EHAQSSLSAAVHVRAGLAPGEVAVAGGSGGSFWSGRRPVPFGDFLIHASFNN
ACATCTLFQKQSDPFGVDVAFPLSLF LLMHGRGVAARVEAGGLTRMANLLYD
SPATLADVPDGRILVADREHNFITVPGCLVENIKSTYLNKLTIVVHGVVSKALPR
STVKVTPQAEFVDLAWLSGGAGGGGCVFVGGIGQPCPADARLYVALTYEAGPR
FTFFOSSRGHCQIMNLIIRIYSPSTMHRYAVVQPLHIEELTFCAVACLGTSATDGMW
RSAFNRYGSSSLPVVEIDSFYSNVSDMEVIL"
complement(75322)
/notes="TATA: TATTAG before BOLF1"
75819
/notes="TATA: TATATAAG"
75838
/notes="polyA signal: AATAAA"
complement(76126)
/notes="polyA signal: AATAAA"
76169
/notes="TATA: TACATAT BO-R2 early promoter before BOLF2,
gives 2.8kb RNA"
complement(76300)
/notes="polyA signal: AATAAA"
76407..78887
/notes="BOLF2 early reading frame, 2 NXT/S. Homology HSV
140K ribonucleotide reductase (Gibson et al, 1984) and RF
19 VZV; NCBI gi: 832980"
/codon_start=1
/translating="MATTSHVEHLLSKLIDELVKANSDEADVIAGRLHRIKAES
VTHVVAEYLEVSQKFEDEFFQMRDELTRVSFAQSPAYERIVSSGYLSALRYD
TYLYVRSQKQESQYHRLAGFCASTTCLYAGLRAALQARPEIEMDFEYDFE
HLTSQTVCCSTPFMRFAGVNSTLASCLITLTPDLSSEWDVTQALYRHLGRYLFQARGV

```

promoter

misc_feature

promoter

CDS

promoter

promoter

misc_feature

misc_feature

promoter

misc_feature

CDS

```

GVGVTCAGCGKHISLIARMINSVEYHNYCKRPVSVAAMFWSHQIFKELETKLP
ENERPCGFTGLFVPELFFKLFERDTPMSDWYLFDPKQAGDLERLYGEEFREYERILV
TACKCCGRVSTKSIWFSIVNCVAKGSPFILLKEACNHFVRDLOGEAMNANICAEV
LQPSRKSVATCNLANICLPCINWNPALVRAQADTDQDELLALPLRSVLTPEGAV
GDGFSIARIDATQCATVACVACSLQAGSPTYDSRDMAGSLGVQGLADVFADLGWOYT
DPPSRJAKKEIFEHYFTALCTSLGLHTRKIPFGKOSKYAGWFHMHDMACTDLS
IPREIWSLERIVRDGLNSQFIALMTSCGAQVTCGSDAFYFFYANASTKYTNKEE
ALPBNRSHRHLDRALINLVGGRVSCIPLEARLQRYLRQTAFTDINOEDLIQMSRD
RAPFVQSQSHSLFREDDAARASTIANILVRSVELGLKTTIMYCYRIEKAADLGMWEC
KASAAVLPREEQNERSAEQMPRMEPAQVAGPYDINSGPCGCGCWCPGCGLEV
CYKROLESSEDDLLETGDFTERACESQ"
/notes="BAM: Bam H1 0/a"
77835
misc_feature
/notes="BAM: Bam H1 0/a"
78804
promoter
/notes="TATA: TATAAGT Ba-R1 early promoter before BMRf1,
gives 3.5kb RNA"
78883
misc_feature
/notes="polyA signal: AATAAA, end of 3.9kb late RNA from
75017 and 2.8kb early RNA from 76169"
78900..79808
misc_feature
/notes="polyA signal: AATAAA"
/notes="BMRf1 early reading frame, homologous to HSV 38K
ribonucleotide reductase (Gibson et al, 1984) and RF 18
VZV; NCBI gi: 832981"
/codon_start=1
/translation="MSKLLVYRDHEGACLVETHNRNFWAAHIVLTKDCCGLKLINE
RDLEFYKLTFFLAAEKLVNFNIDELVTSFESHIDHYTEQKAMENVHGETYANIL
NMLFDGDAANNAVAAEIMADEALQAKISMLRQKAAVATLPEKILVFLITGEGFIS
SFYSIALVRGLMPGICLANNYSRDELLHTRAASLLNSWNTAKADPRATWQELF
RTAVEVEFTAEFACGCVTLVDRAIKQFLLEATADRIILGDIQAPLYGTPTPKDCPLT
YMTSIKQNFEEQESSDYTMLVDVDDL"
complement (79495)
promoter
/notes="TATA: TATAACA"
79537
misc_feature
/notes="BAM: Bam H1 a/M"
79840
promoter
/notes="TATA: CATAAAT BM-R1 early promoter before BMRf1,
gives 2.5kb RNA"
79899..81113
CDS
/notes="BMRf1 early reading frame. Early antigen protein
recognised by R3 monoclonal (Pearson et al 1983; Cho et
al, 1985a); NCBI gi: 832982"
/codon_start=1
/translation="METQTILRFKTKALAVLSKCYDHAQTHLKGGLVQNLVSNYGG
PRLAAVANAGTAGLISFEVSDAVAEQNHQSPEEAPAVSFNRLAYGRTCVLGKELF
GSAVEQASLQPKYRPGGSRPEVKLTXYDDKVSKSHHTCALMPYSPADSLRNEQ
MIQOVLISQPKQKARQSGGVKVTLPNDLVVVTYTSGEACTLIDVKPLSVGP
YEAFTCPVAKAQDVGAZEAHVVCVAADSAAALSICRIPAVSPVILREYSGLIAV
AGLLTSAGDPLDLVSLFNFHASEAAASTASEPEDKSPRVQPLGTGIAQQRPHTVSP
SPSPPPPTPTWESPAPETTPSPATPSPSHSNTALERPILAVQLARKRTSSSEARQKQK
PKVKQAENPLI"
80779
promoter
/notes="TATA: TATTAA BM-R2 late promoter before BMRf2"
complement (80782)
misc_feature
/notes="polyA signal: AATAAA"
80832
promoter
/notes="TATA: GATAAAA, possible promoter for 1.4kb late RNA
encoding BMRf2"
81118..82191
CDS
/notes="BMRf2 early reading frame; NCBI gi: 832983"
/codon_start=1
/translation="MFSCKQHLISGACVFCIGLLASTPTWCFVFNALLSLEIFSPWQ
THYVRLGPTACLMAVLMTLPKHAHRAVRAVTPAIMLINTASALIFFSLIRYSTWISA

```

```

PCLFLANILPLLCMLPRIAIEIVYICPAIHQRFELGELLACTIFALSIVSVSALESVA
FMSPPFFIFLAGSGSLAGARRNQIYTGLERRRSIFCARGDHSVASIKETILKRCPDLL
LAISALIVIVCVMIVLVHVAEVEFGLSRVLPLELGGMASGCLYLGHSSITACVAT
LCTLSVVVYFLEHETLGPLKTVLIFISIFYFSYFSGVAALSAMRYKIKKFNCPGLVHL
RVVVMCCFVFTCEYLLVTFIKS"
81751
promoter
/notes="TATA: CATAAAT"
82180
misc_feature
/notes="polyA signal: AATAAA, end of 3.5kb early RNA from
79804, 2.5kb early RNA from 79840 and 1.4kb late RNA"
complement (82311)
promoter
/notes="TATA: CATAAAT"
82319..82461
repeat_region
/notes="2x71bp repeats"
complement (82746..84122)
CDS
/notes="BMLF1 early reading frame. Diffuse early antigen
(Cho et al, 1985b). Also homologous to RF 4 VZV and IE63
of HSV (BSIF2 + BMLF1) is also called ER2
(Chevallier-Greco et al, 1986). General transactivator of
transcription (Lieberman et al, 1986)."
complement (82747)
misc_feature
/notes="polyA signal: AATAAA"
83640..83729
repeat_region
/notes="10x9bp repeats"
complement (84122)
misc_feature
/notes="ACCEPT: CTCCTCTCCAG acceptor in spliced form of
BMLF1 RNA"
complement (84227)
misc_feature
/notes="DONOR: CAGGTAAGA donor in spliced form of BMLF1
RNA"
complement (84229..84288)
CDS
/notes="BSIF2 early reading frame in 5' exon of spliced RNA
encoding BMLF1; NCBI gi: 832984"
/codon_start=1
/translation="WPSQRLSRTSSISSNEDPA"
84233
misc_feature
/notes="BAM: Bam H1 M/S"
complement (84257..86881)
CDS
/notes="BSIF1 reading frame, homologous to RF 6 VZV; NCBI
gi: 832985"
/codon_start=1
/translation="MSAPVVKALVANSNTDIAEILDAILSRDEGEFRFLCCHNASP
LHHVAGSLVELQLHLPKRLTSQSRGGLVTLHLPAEAPFFLRGLTFLADRSLSTYL
DRAGALRSJLPVLELTLISAKQPOGARGVAMLRPKIVGCLIRYRVNISARMFIS
TFGSHQAQFVIVTAAYVFWGIPCTIETLAHLELTFESQSLAAVTSIAELGEVGS
SWAQEQEAFANFAHEKLRDSREIRAVARTIDAYRGRLPLASADLVRYVYLAHQCF
NEGTFKYSQJLTSMTGECILPGGGVVLPSLLDRGFAEHMRTYFTRETYLAEHVRVQOL
KIRMEPPAPYTMDDPDDGLMARAGLSVDARELVAELARHMADEGCTPTPTLQGFCL
LAQATCGCQWNPKEQFLPPTVLRVQRLPVELCHFDADRHVYFVMTAADPFSSHAERV
PTNCRPLDITRALSYTPVYYSQNSLSEQLFVSRHEYFNRPVPCNLVLDLIDKI
KAPNSLEETIDLCRTVRREVLRMLRGLRGPVRAHPVFFKACPPADPNMEDVLPF
CCTCKLGRFVITPLPRCHAIYCTSAVQGVSVLQKLMGLTACLRMRBHKIKEIAPL
FDSYGHAGCRILPDKYVRDGGLSRQLRFVCHPEEDKHYSYKNAINTQNLHH
SLRWGWPAPKTCYHIADDRDYLIORTRETPPTVENVCAMTEGLGLDGLVAVWSSC
IMPSLMJLATVAPEDKEPQFLHVTFEQTPNLVQVCHARGNFACLRHTRHRSKNV
RFVILVYTSQAITVTMSQCFAGCCGANOPTAHFSISVPASRLINRAESQSSTTSQ
LARRDRQDGSFETPN"
complement (84356)
promoter
/notes="TATA: CATAAAT before BSIF2 and BMLF1. Two RNAs
start here; one is spliced and the other is unspliced,
both traverse BMLF1."
86882
promoter
/notes="TATA: TATTAA BS-R1 late promoter before BSRf1"
86924..87580
CDS

```

```

/misc_feature
/translation="MAFYLPDMSCCGLWLFGRPNRYVSOLPEETFECPDRWRAEID
LG1PPGVQVQGLLNQNTGMSLRVYLLVAOANSITDHLKRPDAVRVPSRCRGVDAQ
VAKLEAVRSVITWMTISLAVSGITMDENGKALKLDKQAGDSLALMEMEKVATALKRQ
TGAAQETSIAVSVSVTAPASAPFINSAPFINSAPFINSAPFINSAPFINSAPFINSAPF
"
complement (87134)
/notes="polya signal: AATAAA"
87599
/notes="polya signal: AATAAA"
complement (87613)
/notes="polya signal: AATAAA, end 1.0kb early RNA from
BL13"
complement (87638..88474)
/notes="BL1F3 early reading frame (BL1F2 in Baer et al,
1984). Homologous to RF 8 VZV and dUTPase HSV. NCBI gi:
832987"
/codon_start=1
/translation="MEACPHIRYAFQNDKLLLOQASVGRLLTVNKTILLRPMKTTTV
DGLIYARPEEGHGLMIGSTRPVTSVHGIIDPGYTGELRLIQNQRNYSNLTLPSEL
KHLAFAYATPQMEDKGPINHPQYPDGVLDSLKDLALPHTVTSVTLVTPPPS
IPIHRTIFGSGLAGMOGLLVKPCRRRGCVNLSLNFSDQTVFLNKYRRCQLVYLH
KPHLTSFYSYSDGADVLGRSLFRWASCTFEVPSLWAGDSGLSEALGRGRGFGSS
CQ"
87650
/notes="BAM: Bam H1 S/L"
88507
/notes="TATA: TATATAT BL-R1 late promoter before BLR1,
gives 1.0kb late RNA"
88511
/notes="TATA: TATAAGA"
complement (88514)
/notes="TATA: TATATAT BL-L3 early promoter before BL1F3,
gives 1.0kb early RNA"
88547..88855
/notes="BLR1 late reading frame; NCBI gi: 832988"
/codon_start=1
/translation="MGKVLKPKFAKAVPLLEFLAATWLLTGVLPAGASSPTNAAASIT
EAQDQFYSYTONADTFSSLSFASIWALLTLVLVLIASAIYLMYCFNKFWNTLLTD
"
88863
/notes="TATA: TATTATA BL-R2 late promoter before BLR2,
gives 0.6kb late RNA"
88925..89413
/notes="BLR2 late reading frame, 2 NXS/T; NCBI gi:
832989"
/codon_start=1
/translation="MSAPRKVRLPSVKAVDMSMEDMAARLARLESNKALKQVLRGC
ACASSTVSPAPVPPPEPTARQREVNITQATGLASQAMKIEDKVRKSVGVGTNRN
EMENILQNLTLRILQVSLMGAGQSPGCEGTPRESNDPNATRRARSRSRGAEKQVQI
SD"
89412
/notes="polya signal: AATAAA, end of 1.0kb and 0.6kb late
RNAs"
complement (89425)
/notes="polya signal: AATAAA, end of 0.7kb early, 2.2kb
late and 2.8kb late RNA"
complement (89430..92153)
/notes="BL1F1a, late reading frame, gp350 membrane antigen,
36 NXT/S (Hummel et al, 1984; Biggin et al, 1984; Beisel
et al, 1985); NCBI gi: 832992"
/codon_start=1
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFFPYPTCNVCTA

```

```

DNNVTINFVGKKHQHLDLDFGQLTPHTKAVQPRGAFGGSNAIHLFLELIGAGEL
ALTMRSKKLP INVTGEEQVSLSDVDFVGTMMCHAEQNPVYLIPETVPYI
KQDCNCSNITAVRAQGLDVLPLSLPTSAODSNFSVKTEMLGNEIDIECIMEDEGEI
SQVLPQKCNKNTCSGESHVSFGGLISLTSVPATPGTGAYSLRLTPRPVSRFLG
NNSILYVYSGNGKASGDYCIQSNIVFDEIPASQDME TTTDITYVGNATYSVP
MYTSEDANSPNVTATWAMPNNTEDFCKKWTLSGTPSCGNSISGAFASNRTFDTIT
VSLGTAPKTLIITATNATTTTHKVFESKAPESITTSPTLNTGTFADPNTTGLPS
STHVTNLITAPASTGPTVSTADVTSPAGTISGASVTPSPMDNGTESKADMTS
STSPVTPTPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPT
PNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTP
ATNHLTGTSPTSPVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSP
NNTSHPMLLTSAHP TCGENITQVTPASTISHTVSTSPAPRPGCTTSQASGPNSTST
KQCEVNTKGTTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTP
PTDYGGSSTPRPRNATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNAT
SNLSMLVQWASIAVLITLILLLVMDACAFRRNLSTSHYTTTPPYDDAETVY"
complement (89430..92153)
/notes="BL1F1b, late reading frame gp220 membrane antigen,
spliced form of BL1F1a (Hummel et al, 1984; Biggin et al,
1984; Beisel et al, 1985); NCBI gi: 832991"
/codon_start=1
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFFPYPTCNVCTA
DNNVTINFVGKKHQHLDLDFGQLTPHTKAVQPRGAFGGSNAIHLFLELIGAGEL
ALTMRSKKLP INVTGEEQVSLSDVDFVGTMMCHAEQNPVYLIPETVPYI
KQDCNCSNITAVRAQGLDVLPLSLPTSAODSNFSVKTEMLGNEIDIECIMEDEGEI
SQVLPQKCNKNTCSGESHVSFGGLISLTSVPATPGTGAYSLRLTPRPVSRFLG
NNSILYVYSGNGKASGDYCIQSNIVFDEIPASQDME TTTDITYVGNATYSVP
MYTSEDANSPNVTATWAMPNNTEDFCKKWTLSGTPSCGNSISGAFASNRTFDTIT
VSLGTAPKTLIITATNATTTTHKVFESKAPESITTSPTLNTGTFADPNTTGLPS
STHVTNLITAPASTGPTVSTADVTSPAGTISGASVTPSPMDNGTESKADMTS
STSPVTPTPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPT
PNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTP
ATNHLTGTSPTSPVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSP
NNTSHPMLLTSAHP TCGENITQVTPASTISHTVSTSPAPRPGCTTSQASGPNSTST
KQCEVNTKGTTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTP
PTDYGGSSTPRPRNATVTPPNATSPATVTPPNATSPATVTPPNATSPATVTPPNAT
SNLSMLVQWASIAVLITLILLLVMDACAFRRNLSTSHYTTTPPYDDAETVY"
complement (89434)
/notes="TATA: TATAAG"
complement (89567..90013)
/notes="BL1F2 early reading frame (BL1F3 in Baer et al,
1984); NCBI gi: 832990"
/codon_start=1
/translation="MCPVHQHPAQAPPAKQALETVHPQNRGLMSPKARPPKMQR
RRPPVAKRRFRSPQOQVERILPPVESTPDQMEPQQVQSPQITAVIQLRQDRDTH
RPPIYLALLANGCPAGLIRAHRLPQPKPCQSRQRPSPDSQTSPC"
complement (90051)
/notes="TATA: TATAACA BL-L2 early promoter before BL1F2,
gives 0.7kb early RNA"
complement (90062..90652)
/notes="intervening sequence in gp220 gene"
90177..90639
/notes="21 copies of 21bp approximate repeat"
complement (92192)
/notes="TATA: TATTAAA BL-L1 late promoter before BL1F1a,b.
Gives 2.8 and 2.2kb late RNAs"
92238..92581
/notes="Exon in Bodescot et al (1986) RNA (spliced from
20763 to 92670)"
92243..92581
/notes="BLR1F3 reading frame; NCBI gi: 832993"
/codon_start=1
/translation="MDKOREGCPALDNDMEEEVPTSVQEQVSAGDWNVLIELSDS
SSEKAEADHLEPAPQKGTGRKRVDDHAGGAPAPMLPQPQDLP GREATLRREFLDLRL
TLIQAIGAAAT"

```

```
CDS
92670..95162
/notes="BERF1 frame, homology with BERF2b and BERF4. A
fusion of BLRF3 with BERF1 encodes EBNA-3a, latent cycle
gene. (Hennessey et al, 1986, Joab et al, 1987)"
92670..95248
/notes="Exon in (Bodescot et al, 1986) RNA from 92581, to
3' end"
misc_feature
92703
/notes="BAM: Bam HI L/E"
/notes="BAM: Bam HI L/E"
promoter
complement(93161)
/notes="TATA: CATAAAT"
93479
/notes="TATA: TATAAGA"
/notes="TATA: TATAAGA"
promoter
complement(93482)
/notes="TATA: TATAAAT"
94208..94277
/notes="repeat type A"
repeat_region
94281..94306
/notes="repeat type B"
repeat_region
94307..94381
/notes="repeat type C"
repeat_region
94386..94411
/notes="repeat type B"
repeat_region
94412..94489
/notes="repeat type C"
repeat_region
94490..94560
/notes="repeat type A"
repeat_region
94571..94648
/notes="repeat type C"
repeat_region
94649..94719
/notes="repeat type A"
repeat_region
94896..94982
/notes="repeat type D"
repeat_region
94983..95069
/notes="repeat type D"
misc_feature
95221
/notes="polyA signal: AATAAA"
complement(95272)
misc_feature
/notes="polyA signal: AATAAA"
95353..95724
CDS
/notes="BERF2a reading frame; NCBI gi: 832994"
/codon_start=1
/translation="MKKAWLSRAQQADAGCAGSEDDPPYDQGNVTQVGSEPISEI
GPFELSAASEDDPQSGPVEENLDAAREEEEPHEQHGDDPLDVHTQPRFYDVP
TOAPVIQLVHAVYDSMLVRGT"
95725..98244
CDS
/notes="BERF2b frame, homology with BERF1 and BERF4. BERF2a
and BERF2b are spliced together to make EBNA3B (EBNA4A)
latent protein."
complement(95819)
misc_feature
/notes="polyA signal: AATAAA"
95853
promoter
complement(95853)
/notes="TATA: TATAAAT"
misc_feature
complement(96276)
/notes="polyA signal: AATAAA"
97522..97698
repeat_region
/notes="3x60bp repeat"
98323..98769
CDS
/notes="BERF3 reading frame; NCBI gi: 832995"
/codon_start=1
/translation="WTILDEVEILHIFRPTMESFEQGSRSQSPDNERGDNVQTGGEH
DQDPGCPSSGASERLPEESYRQQPWQSGSDENCMQRIRRRRRRALSCH
LLQEDNVFWPLPHDITPTTARNIRDAACRAVKVSMPLTGMCAI"
98364..98730
mRNA
```

```
misc_feature
98731
/notes="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
/notes="DONOR: AAGGTGAGT donor"
98805..99050
mRNA
/notes="Exon in T4 cDNA (Bodescot et al 1986) . 99050 is not
the end of the RNA."
98805..101420
CDS
/notes="BERF4 frame, homology with BERF1 and BERF2b. BERF3
and BERF4 are spliced together to make the EBNA3C (EBNA
4B) latent protein."
99126..102118
misc_feature
/notes="DEL: Deletion in Raji"
99443
/notes="TATA: CATAAAA"
100104
misc_feature
/notes="DONOR: ACCGTGAGT possible donor before repeat."
100122..100304
repeat_region
/notes="10 x 15bp repeat"
100528
misc_feature
/notes="DONOR: CTGCTAAGG possible donor"
100613
misc_feature
/notes="BAM: Bam HI E/e1"
100665..100781
repeat_region
/notes="3x39bp repeat"
complement(100860)
promoter
/notes="TATA: TATAACA"
100919
misc_feature
/notes="BAM: Bam HI e1/e2"
101426
misc_feature
/notes="BAM: Bam HI e2/e3"
complement(101445..102116)
CDS
/notes="B2LF2 reading frame 3x NXT/S. 2.5Kb late RNA
traverses B2LF2, ends unknown. NCBI gi: 832996"
/codon_start=1
/translation="WVSFKQVRVPLFTAIALIVILLVFLPVRVGGRRVAAAALTW
VPKPNVEWVPDPPPPNFNKTAQCYGDKVKLPHWTFILHTFQVQNTKANCYIC
NTRYETSYKCCFYTKKKHWMGCAQCAELPCTTFYGPDPDLPVWTRNNAIE
SLWGVYRVGEGNWTSLDGGTFKVIQIFGSHCTVTSKFSTVPVSHHESFLKPCLCVS
QBSNS"
101690
promoter
/notes="TATA: CATAAAA"
101765
misc_feature
/notes="polyA signal: AATAAA"
complement(101786)
promoter
/notes="TATA: TATAAG"
101947
misc_feature
/notes="BAM: Bam HI e3/2"
complement(102098)
misc_feature
/notes="DONOR: CAGCTGAGG possible donor"
complement(102126..102341)
mRNA
/notes="3' terminal exon of 0.9kb and 2.8kb early RNAs"
102153
promoter
/notes="TATA: TATTAAT"
complement(102156)
misc_feature
/notes="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs
encoding B2LF1 and BRFL1"
complement(102160)
promoter
/notes="TATA: TATTAAT"
complement(102213..103155)
CDS
/notes="B2LF1 reading frame, modified from Baer et al,
1984. Has two splices within frame. 2xNXT/S. Immediate
early gene which disrupts latency (Countryman and Miller,
1985), called EB1 by Chevallier-Greco et al, 1986 and
ZEBRA by Miller."
```

```

promoter      complement(102380)
              /note="TATA: CATAAAT"
promoter      102415
              /note="TATA:TATATAC"
promoter      complement(102420)
              /note="TATA: TATATAC"
mRNA          complement(102426..102530)
              /note="Exon of 0.9kb and 2.8kb early RNAs"
misc_feature  complement(102504)
              /note="polyA signal: AATAAA, apparently not functional"
repeat_region 102581..102652
              /note="semi-repetitive sequence, homologous to human c-fos
              3' sequence"
mRNA          complement(102655..103194)
              /note="First exon of 0.9kb early RNA encoding B2LF1"
misc_feature  complement(102918)
              /note="splice acceptor used in RZ fusion gene (Sargeant)"
promoter      complement(103231)
              /note="TATA: TTTAAA of B2L1 immediate early promoter gives
              0.9kb RNA"
misc_feature  complement(103256..103311)
              /note="Upstream of B2L1, homology to 106243 to 106188"
CDS           complement(103366..105183)
              /note="BRLF1 reading frame, (immediate?) early gene, acts
              as transcription activator. NCBI gi: 832998"
              /codon_start=1
              /translation="MRPKKGIEDFLRLTPEIKKQLGSLVSDYCNVINKEFTAGSVEI
              TLASYIKCAFINEAKAGREWGGLMATINCNFWAILNNRVRRRAENAGNACSIAT
              CPTVMRYVLDHLVTDTRFTQAPSNRMVPIATICTAMVKLLKHSRVAYTYSKVLGV
              DRAAIMAGKGVVHELNRMWEKGLSSKFAFCNWFVTPVLEEMFQTNWSKGTGHLT
              DVVKVRALIKITLPRASYISHAQGSYSGVLPACLLSTSKAVETPILVSGADRME
              ELMGNDGASHIARYSESQGFHAFTDELESIPSPMP LKPCQASADCGDSSSSDS
              GNSDTQSEREAREAPRLAPKSRRTSRNRCQTPCPSNAEPQPHIAVHQESD
              ERIFPHSKPTFLPVRKKGLRSGMFLPKPEAGSAISDVFEQREVQPKRIRP
              FHPGSPWANRPLPASTLPTGTPVHEPVGSLTPAPVQPLDPAPAVTPEASHLEDDP
              DEETSQAVKALREMDVTIPOKEAAICQQMDLSHPPPRGHLDLITTTLESMTDLNL
              DSPLPETNEITDFTINDECLHAMHISTGLSIFDTSLF"
              complement(103453..103462)
              /note="TAATGAAATC sequence"
              103741
              /note="BAM: Bam H1 2/g"
              103816
              /note="BAM: Bam H1 g/R"
              complement(104926..105185)
              /note="exon in RZ fusion gene (Sargeant)"
              complement(104927..104989)
              /note="BRLF2 poss. small 5' exon"
              105016
promoter      /note="TATA: TATAAAT before BRRF1, possible promoter for
              1.1 kb early RNA encoding BRRF1"
              complement(105019)
promoter      /note="TATA: TATAAAT before BRLF2"
              105182..106114
CDS           /note="BRRF1 early reading frame; NCBI gi: 832997"
              /codon_start=1
              /translation="MASSNRGNARPLKSFLEHLYIKHPEYGVGVHLLNTIGVDCDLP
              PSHPLTQAQGLGLIARVLAQVQKHLEDTIVPKLKLKAYLELLSYSPKQDRDI
              AEYLDHLKTRNDLGRDLMALIKRQDRHASVNVLMGSDYTAVSLOYDGLSIG
              MRKVIADVCRSCYASNPSTWATNLSHQLLMACSPSEPCAWRGCFNQLVLTVALCK
              FRFCIYYNIQGSATIATISQLLHLEIKALCSWITSQDMRLRQHSRPLLTLMESVAANQ
              EVTDATITLPCAEYIDLLKTKHVLNCSAMQYK"
              complement(105185)
              /note="ACCEPT: splice acceptor in 2.8kb early RNA encoding
              BRLF1 and RZ fusion gene (Sargeant)"

```

```

promoter      105213
              /note="TATA: CATAAA"
misc_feature  106110
              /note="polyA signal: AATAAA, 3' end of early 1.1kb RNA
              encoding BRRF1"
              complement(106125)
misc_feature  /note="DONOR: CAGGTAAGA possible donor"
              complement(106188..106243)
              /note="Homology to upstream region of B2L1"
promoter      complement(106213)
              /note="TATA: CATAAA"
              106243
promoter      /note="TATA: TATAAAA before BRRF2, possible promoter for
              1.8 kb RNA encoding BRRF2"
              106302..107915
CDS           /note="BRRF2 reading frame; NCBI gi: 832999"
              /codon_start=1
              /translation="MSQQRGSVILVPEHLAAGALTKIAMSDFITGQDVTLSGGMIAVKI
              RDAINOTPGGDOVALSSIFALWNALPTSGRSSRDDLIAPAVALTTAHNLCIGVIP
              GETSKHPTESLIRAITVGLQKLWDSGCCPECIACIQKGLKAIKPLGYEIPRIIPHTK
              QCSVNLNNLHVHLKVLARGVQLAYDARVLTPDFEIPDLDDSDAVFARTLLAALFH
              LNWFLIKDYITQDSMSLKQALSCHMSATGNPLPAAPETLRDYLEAFRNSDNHFYLP
              TTGPLNTQFPPELLGRVWVDSILCAASHVQDVITHGVGAGVPRPFSALPAPSRE
              PQOTCSOLTSGRNESRRNIPGGTSPVPPVCPVSLTASCAKQNRGSGSLHLAK
              PEETSPAVSPVCIAPASRSRKHQCVGTGSOAASFSSVAPVASLSGDLEEEEGS
              RESPSLSKKGGDEEFANLEAODANLEDVQREFSGLRVIGDEDEGSGDECEFSDDL
              SOSDHGDEGGGAVGGGRSLHSLSLVV"
              complement(106385)
              /note="TATA: GATAAAA"
              complement(106973)
              /note="polyA signal: AATAAA"
              complement(107124)
              /note="TATA: GATAAAA"
              107457
              /note="BAM: Bam H1 R/f"
              107565
              /note="BAM: Bam H1 f/K"
              107914
              /note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding
              BRRF2"
              107942
              /note="ACCEPT: splice acceptor for EBNA-1 RNA (from
              98730)"
              107950..109875
              /note="BRRF1 encodes EBNA-1 protein, latent cycle gene.
              NCBI gi: 833000"
              /codon_start=1
              /translation="MSDEGPGTGCNGIGEKGDTSPEGSGSGPQRRGGDNHGRGRG
              RRGRRGGRCFAPGGSGSGFRHRDGVRRPQKRPSCIGCKTGCTGACAGAGGAGG
              AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
              CGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              SPRRPPPPGRPPFFHPVCEADYVEYHOEGGPDCEPDVPPCAIQCGADDPCEGFTGCP
              RQDGGRRKKGFGFKHRQGGSNPKNFENIAGELRALLARSHVTTTDECTWAGVF
              VTGGSKTSLYNLRRTALAIQCRLTPLSLPFGMAPGFGPQCPRLRESIVCYFWFL
              OTHIFAEVVKDAIKDLVMTKPACTCNIRVTVCSFDGVDLPHPFPVMEGAAGCGDG
              DQDEGDEGDEEGEQE"
              108217..108924
              /note="EBNA triplet repeat GGA, GCA, GGC."
              109856
              /note="DONOR: AGCGTCAGG possible donor at end BKRF1"
              109905

```



```
misc_feature /note="TATA: TATTAA before BKR2, possible start for
2.3kb late RNA"
109906 /note="polyA signal: AATAAA"
misc_feature 109937 /note="polyA signal: AATAAA 3' end of EBNA-1 RNA"
CDS 109958..110371 /note="BKR2 reading frame; NCBI gi: 833001"
/codon_start=1
misc_feature /translation="MRACVFLAICLVITFVPLTWGNWAYCCCHVTQLRAQHLLALEN
ISDIYLSNQCOCGFSIASINSRPNQSLVSRKANGLVNVSFFISILKSSSALTG
HLRELITLTLYGSFSEDFLGANLRYAHRGG"
110271
misc_feature /note="DONOR: TCCGTGAGT possible donor at end of BKR2"
110275..111117
CDS /note="BKR3 reading frame, homologous to RF 59 VZV"
111098
misc_feature /note="DONOR: TCGGTGAGA possible donor at end BKR3"
111107..111784
CDS /note="BKR4 reading frame, contains complex repetitive
sequence"
111272
misc_feature /note="DONOR: GACGTGAGT poss.donor before rpt.seq. in
BKR4"
111719
misc_feature /note="polyA signal: AATAAA"
111787
CDS /note="polyA signal: AATAAA : currently unknown which is
3' end of the 2.3kb late and 1.1kb early RNAs"
complement(111830..114259)
/misc_feature /note="BBLF4 early reading frame, very good homology to
RF55 VZV; NCBI gi: 833003"
/codon_start=1
/translation="MAEPRAPALSTFTMLNWTSDASVRIIVRRIGTLARRVQQLP
DMETSPFDELPSELPFSAVITGTAGACSTSVSCLHHTMDCLVGTATTAAQ
NLSTQIRACPTVYSAGFKSRHINMTORVSSHGSRSTDALELQRRDLAKWPLSD
IAEFRRTRKGLYSGVSGPAFEAFVLDHMQQLMTNVIIVDEAGTSLWHILTAWFC
YWFNNALRTPLVRRRCIPICVCGSPGTDAFQSSFSHETQVNRKRCMDNILFTLVG
NPAATYVDVARNWALFNNKRCQDVQFCHLMKTLEYGLESPDILAYVDFVPRAA
IMDPAQYGVNTRFLSHAENVKTEFLTTHATLKTAGAGRAAGTGGGGGVTFCTPVE
CEYFLDPLAQTKTLVGLPCLTAHTWLQKNYARLNTYSGAQDMVPVPTQEDDERVKV
TNVTVYKHSVSNNCKTKSCYCTGTFGDMTLEADSFEAHGHEQPEVYVFLA
ERSPAAVYARVLAPPANASAPLCSLNITYNDLFRQCLDVAVRGGRFDLPFC
TFNNMLIRDNLEFSDPELHRLGLDYASTTENYTLGGYHLNVFTGIRGKQOPDAG
SSRMPRLWDEAGFVCCLEHNWKLYETIEDKSLNLCSDIRYGISSKLMTIKAQG
LSLNKVAFCFSGHNRNFKPGRVYVALLSRASHNVCVMDRNP LSEMITGCPNASCYVD
ALKNSRALLY"
misc_feature complement(111830)
promoter /note="polyA signal: AATAAA"
112471
promoter /note="TATA: TATATAT"
complement(112476)
/misc_feature /note="TATA: TATATAA"
112620
promoter /note="BAM: Bam H1 K/B"
113876
promoter /note="TATA: TATTAT before BBRF1"
complement(113885)
CDS /note="TATA: CATAAAT"
114204..116045
/note="BBRF1 late reading frame, homologous to RF 54 VZV;
NCBI gi: 833002"
/codon_start=1
```

```
/translation="MFNNVDESASGALGSSAIPVHPPTASVRLFEILQKAYVQGG
TIYANLRNPGVFSRQVFTHLFRAISHCTYDDVLHDMNKFEACIQKRWPSDDSCASRE
RESTFESKSTTMKLTVRDLTLTNIVRVLHRSRVSLSVERYVDWICATGWPAAVKKPTQ
ELHSKINSLRDRCVCRELGHRTIRSIGTEYATKEIIESINSITFYQFTEVTEIYL
PRSEYVAYYCCRRIRLHVLPFAIFACTYFDSVQRLYNIFMCTYRTEHAKICQL
LNTAPLKAIVGHGGRMYKDI LAHLEONSQKDPKPKELLALLVKLENKTTISGVTDVY
EEFITDASNVLDRNRRLFCQGETAAQGLKKNYSNVTWKCLTDQINEQFDQINGLEKE
RELYKKIRMSQVLOASLGP GGNPAASAPAAVAASASVDILTGSTASAEIKLFNS
PASIGARVSGHNEISINFSVQYIPPSRGMWTKDLTELMESELFNTFKLTPVVDNQQG
RLYVYSSDTSISILIGPFTLYVAELSPVELVTDVATLGIVEIIDLRYSSRLAIYIE
DUGRYCOPASATGGGHIQAPFARGSTPDHAKSKPAROPPPGAGS"
115843..116781
CDS /note="BBRF2 late reading frame, homologous to RF 53 VZV"
complement(116683)
promoter /note="TATA: GATAAAA"
misc_feature complement(116696)
/note="polyA signal: AATAAA"
CDS complement(116784..117386)
/note="BBLF3 early reading frame, spliced to BBLF2. BBLF3
contains a consensus nucleotide binding site"
116785
misc_feature /note="polyA signal: AATAAA"
complement(117386..117515)
intron /note="intron spliced out in RNA linking BBLF2 and BBLF3"
CDS complement(117515..119080)
/note="BBLF2 early reading frame, spliced to BBLF3"
118981
promoter /note="TATA: TATAAAA BBR1 late promoter before BBRF3"
119067
promoter /note="TATA: TTTAAAA BBR2 late promoter ?"
119098
promoter /note="TATA: TATTAA BBR3 late promoter before BBRF3"
119108
misc_feature /note="DONOR: AAGTCGAT possible donor"
119137..120354
CDS /note="BBRF3 late reading frame; NCBI gi: 833004"
/codon_start=1
/translation="MKSSKNDTFVYRWKTLVTVFVFNVSAPVITAMFPNLGYPC
YFNALVDYCALMTNYNLAHLPTLYLEPPENFVITLVFIADCVAFIYACEVAL
IKARKNSGLTDLASWVSAGVSTVIFLAIKLMSIQVFIQVLSYKHVLSAFVFLH
FLASVLAHACVTRFSPVWYKQADNSIPQDTFLAMVYIKRPVNTNLYLGCLALETL
VTSLSVFLALGNSFYEWGDMVLGAVNLELILP IFWYILTEWLASFLRHNFGYCGM
FIASILLIPLVRYENFVSAKLHTTVAINVAIIPILCSVAMLIRICRIFKSRQGTD
YVPVSETVELESEPRPRSRTPSPGRNRRSSSTSSSSSRSTRQRVPVSTQALVSV
LPMTTDSEETFP"
120260
misc_feature /note="ACCEPT: ATCTTCTCCAGT possible acceptor"
120358
misc_feature /note="polyA signal: AATAAA"
CDS complement(120747..120974)
/note="BBLF1 late reading frame, possibly homologous to RF
49 VZV; NCBI gi: 833005"
/codon_start=1
/translation="MGALWSLCRRVNSIGVDGGGINLYNDYEEFNLETTKLIAAEE
GRACETNEGLEVEDSENDLFLPNKPN"
misc_feature complement(120764)
/note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
early, 3.0kb early RNAs"
CDS complement(120929..122341)
/note="BGLF5 early reading frame, homologous to RF 48 VZV
and alkaline exonuclease of HSV; NCBI gi: 833006"
/codon_start=1
/translation="MADVDELEDPEEMETSYTFAFRLSPFETAFVRLDRPPQPMAM
REVVYLCCLKQIQEFSGETGFCDFVSSLIQVENDSKDGSILKSYMGLQAEATDQRVL
```

CSVSNTRGSGENLMDILANGIISSSKLSTIKNGPKTKPEPAPISTNHYFCGPA
 FGLRCDTVDKIVCKLICDASNRQFGFMISPTDGGIFGVLIDLCVNVESQDFILFT
 DRSCIVEIKREYLFESKEEDPIYPSTALYKRPCKRSFIREINSIRPTEVEVPDG
 RLPSEGVLITQDEANWIKDVRKKLCPGHDLVADSIAANRGVESMLYMTDPSNAG
 RIGIKDRVPVNIENFRNNFYQVLLQYKIVGDIYVHRSKGKRGDCSPRNIIVTAF
 RKSPIDPATCTGLSDLLDASVEIPVAVLTPVPLDSVIRKTLSTAAGSKAYADN
 TFDTPWPSPGLEADDESTP"
 complement (121331)
 /note="TATA: TATTAA BBL1 late promoter before BBL1"
 /codon_start=1
 121669
 /note="TATA: CATAAAT"
 121697
 /note="TATA: TATAAG"
 121772
 /note="TATA: CATAAAG"
 122313
 /note="BAM: Bam H1 B/C"
 complement (122328..123692)
 /note="BGLF4 early reading frame, homologous to RF 47 VZV"
 complement (123506)
 /note="DONOR: AAGGTGACT possible donor"
 complement (123941..124939)
 /note="BGLF3 reading frame; NCBI gi: 833008"
 /codon_start=1
 /translation="MFNAVADMPDDPMIARRYQCILEALEACQDTPQKLVETPL
 KSFLLVSNILPQDNREPMHARSGRAVEDDYDESSIALELLPINRPLEEPCQGGW
 SSRMEPSPEMGCLGFEVDGDLIRIALANKNDEVIGQALQILANSQWTSIVPEDP
 LPMWALYCFRSHCEERHCYAAARGKRPILPTATVPCANIEAFLAHLTCVYA
 LYLDVRMKGEDIAPPDFVSLRANKAKOILCLOEPFCITRVCLLCILLHKNLNAQYK
 RPVDYDCLILGAEARYWYDVGNYREASTGTGTVLYPTDLSIVADMVTEDE"
 complement (1241117)
 /note="TATA: TATAAAA"
 complement (124219)
 /note="polyA signal: AATAAA"
 124938..125915
 /note="BGRF1 reading frame, homologous to RF 45 VZV and
 spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
 Northern blots in BGRF1 detect 2.7, 2.6, 2.1kb late and
 1.9kb early RNAs. 2.6, 2.1kb RNAs very weak. NCBI gi:
 833007"
 /codon_start=1
 /translation="MLYASQRGLTENRNALQDSTTGCGLGAETPSTMTGAKSDR
 WAMPVGTIRASNLCPMLRAYCHYGRPVFADESILPMFGASPALHTPVQVQCL
 LPELRTLQRLPPNLEDEALTEFTKSVSSARAILEDPNFLEMREFVTSIASFLSG
 QYKHPARLEAFQKQVHLHSFYFLISIKSLTIDTFMDFIQSAPGLEEMTEKLIHFK
 QKASVFLIPRHHGTWIVVAIISILSNLNVQICVVAHQKHVASVFTIIDIITLKS
 FDSKRVENKETSITITFRHSKISSTVMCATCNKVRDVSVLGNCA"
 complement (125113)
 /note="TATA: TATAAAT before BGLF3"
 complement (125484)
 /note="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
 late, 3.0kb late and 3.7kb early RNAs"
 complement (125863..126873)
 /note="BGLF2 late reading frame, poor homology to RF 44
 VZV; NCBI gi: 833009"
 /codon_start=1
 /translation="MASAANSRQLRKFLNKECLMWLSDASTPQMKVYTTATVASV
 YVQIAGPPKTYNNVTLVIAKPKKPTVYVYINGTLATVARPEVLFKVAQGPHSIT
 LMYGVFSDVAGVAVRPNWPTCTDLTATVFTTSTAVKVEELQDITSELIIP
 LGRGAWYAEALYMFVNMGMALCCPNPPTSLTFHINLLTRCDNGECVTCYAGA
 HVNILRGWTTDDSPGSGTCCPLPCALNDNDYVPIIGHRALGLMFKPEDAPFVGL
 RFPNPKMHPDMRSVILQCVLANGKEVCTAQPTLLRFSDLYSRAMLNQCIVLRQVHL
 SY"
 126277

CDS
 /note="TATA: GATAAAA"
 complement (126851..128374)
 /note="BGLF1 late reading frame; NCBI gi: 833010"
 /codon_start=1
 /translation="MDVHIDNQVLSGLGTPLLVHLFVDPVTMAELCPNRPVNCCEGAWC
 OTLFSDRTGLTRCVFAARGMLPGRPSHRGTFTSVVYCGEGLPELNPFPVHAALRE
 YDEGGLVEIQIYIISLFEFAGKXALTDGHLIREASGVQESAAAMQPIFIDPGRPAGG
 IEMFVAAQVHPKTYDIAQIILLETENRGEQRLHAGSPALCLGLRLRGAETK
 AAATSVSKHHPALENPNIRSGAGCGGGRAGTGTGVGSGALSRSVPVSFSTRR
 AIRESBALVRGIAHIFSPHALYVTVYTPSLAQGRHLRMTAVTHASPDIDAEVSIIGA
 PEREFRLISVALRISAFREKLAMQAWTAQEIPIVPIYTSIRYKNSDLIREAFFT
 VQTRVSWESQKATISNAPKTPDACLWIDSHPLYEEGASGKGVDSRPPGGIIVGAA
 SQLVALGTGCHGVHLATTSQQAFLVLPGGFEVVKQIALTPEERGVIILARHGIRREQ"
 complement (126929)
 /note="TATA: TATTAAA EEL8 late promoter before BGLF2,
 gives 1.6kb late RNA"
 complement (127237)
 /note="TATA: TATATAA, potential promoter for 1.8kb late
 RNA"
 128029
 /note="polyA signal: AATAAA"
 complement (128344..129021)
 /note="BGLF4 early reading frame; NCBI gi: 833011"
 /codon_start=1
 /translation="MSDQRLSLPRGEGTDEPNRHLCSYKLEFHLPLPESMASVF
 AWCQCGEYHVCDGSSSECTLIEHGVCAITGNYGPHFQPALRPWTEIRQDTQDRD
 KWEPQVQGLVTVVNNLHYFINENISGVSEALFDOEGALRPHIPALVSFVPCCL
 MLFRGASGEKVDVVLISYIIVHSIYSQKTVYCALLFSTRNKRKYDAVAKRRELWM
 SLITTKC"
 complement (128432)
 /note="TATA: TATTATA before BGLF1, potential promoter for
 3.0kb late RNA"
 128848
 /note="BAM: Bam H1 G/D"
 complement (129054)
 /note="TATA: TATTTCG before BDLF4, potential promoter for
 3.7kb early RNA"
 129188..130348
 /note="BDRF1 reading frame, homologous to RF 42 VZV and
 spliced gene in HSV (Costa et al, 1985). Spliced from
 BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
 1.9kb early RNAs. Possibly also 1.8kb early RNA."
 129374
 /note="TATA: TATAAGC"
 complement (129377)
 /note="TATA: TATAAAG"
 129413
 /note="DONOR: GTGGTAAGT possible donor"
 130347
 /note="polyA signal: ATTTAA"
 complement (130359)
 /note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
 2.3kb late RNA and 3.2kb late RNA"
 complement (130362..131066)
 /note="BDLF3 late reading frame 9kXNT/S; NCBI gi: 833012"
 /codon_start=1
 /translation="MAHAKAKAGVAMILICETSLTWTSSGSSTASAGNVGTGAVT
 TSPSASGSPSTNQSLTITTSAPITTTAILSTNTTITFTGTTVPVPTTSNSTINV
 TKTVTAQNIATGACTGTGTCVTSNVTSSSTSATRITNATITLAPTLSSKGSNA
 TKTLTALPTVPPDERQPSGLPLMTLVFVGLTFLMLILIFAAGLMSAKNRKLDLAL
 LNAVTRDPSLYKGLV"
 complement (131104)
 /note="TATA: TATATAA EEL4 late promoter before BDLF3,
 gives 0.9kb late RNA"

```

CDS
complement(131127..132389)
/notes="BDLF2 late reading frame; NCBI gi: 833013"
/codon_start=1
/translation="WDQQAIVEHGVTVSHITISREEDGVVHRRVVASGERVVEFYKAP
APRREGRASTFHDFTVPAANAAPGEPEPEHPMP IHANGGETKNTDQDNQOT
TRTRTNAAKEARTAEWDDTMASGCGORAPISADLLSSLTGRMAAMAPSMKSEVVC
GERMRKEDVGEAETLAEPFRCFMLSFVYIYCCYLAFLALLAGFNPLFLPSPMP
VGAKLRGKGRDGVPLSTGYCNPFCVKYTLIPAVV INNVYYPNNVDSHGSGFFE
AAALHVAALFESGCPNIAQVNRNRNFNTRASGRVERRLVQDMQVILASAVVMHHH
CHVETVYVGVGPEEGTPTPCFKDVIARPSLVNTCTAPLKTSVKGNWSGAAGM
KRKQCRDLRDSFPAYLEEVMYVMVQ"
132266
/notes="TATA: TATAAAA"
complement(132400..133305)
/notes="BDLF1 late reading frame, poor homology to RF 41
VZV; NCBI gi: 833014"
/codon_start=1
/translation="MDLKVWVSSRLYTDIEAKMQORIGCIPLASTHGTQNVQGLG
LGQVSLQVTPDVSNVNYLSDCTLAVLDEVSVDLSLTKVPGQTVAKNKYQPFQ
WHGTGSLWMPVFCREHATVKLESNDVDIVFPMVLPTTIAEVLQKILLNWSRVV
MQAPGNADMIDVHMHGVSYLGHYETALPEVPGPLGLALLDNLISLYFCIMVTLPR
ASMRVIRGLIRHEHDLNLFQEMVPEIARIDLDLSDVADLSRMRVNMVTLQSLAS
LFNLGRLTAAYSQETLTATCWL"
complement(132476)
/notes="TATA: TATTAAA before BDLF2, likely promoter for
2.3kb late RNA"
promoter
/notes="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
complement(133312)
misc_feature
/notes="TATA: TATTAAA before BDLF2, likely promoter for
2.3kb late RNA"
CDS
complement(133321..137466)
/notes="BcLF1 late reading frame, homologous to RF 40 VZV
and major capsid protein of HSV; NCBI gi: 833015"
/codon_start=1
/translation="MASNEGVRNPPYLYTDADLLSNRQSAEGLHFSDLLVGKD
ARAGTEGLVGVYNAIQYRLETAVALSVNTEFKDLSRMDGKIQFRIISVPTI
AHGGRPSKORTFIVKNCCHKHIISTEMELSMLEILHGISPETVEAYVCAVKT
VASAQFGVDALGERGLNTVSLVKLRHAPPMEILQTLADPTTTERGESKTVKSDLLAM
FKHLLHLEFDLRAENMGSGFQOYVRSLSEMAAVSGESVKGVSYYTKAGGEPV
GGVFIWTDNVLQTLTLEEADNQIMGPSYASFVVRGENLNTAVSYGRVMTFHFH
MARIVDSKAGSTKSDLPAAVAGVEDQRPVISAAYIKLGNHVAVESLQKMYNDIQ
SP1PMDRQWQSYTFEPVGLFMPNPKYTTSAAIKMDLNPQQPLPVEAWIVNKNLLAF
NIQAALVKLCHPLRHPHTLNLNAPAPDRDRETSYLSQRRNHNHNVIVDEFYD
NKYAAPTDLIAKGLPTEDFLHP SNYDLLELHP LYDVIYGRDAGERARRAVHRL
MVGNLPTLAPAFQZARQQGFETATSLAHVQDAVETVQDTATDTAFYVVEA
MIRGFEEKFVMNVPLVSLCINTYERSGRIAFNSFMTKICRHGNNAISKEAYSM
YRKIYGEILIAEQALMLAGSDVGVGVQVYCALLDPLLPVAYTDIFTHLLTVS
DRAPQIIGNEVADTLAQQF IERVGNWDEMAAQFVALYGYRVNGDHDHDFRLHLP
YVDEHADVLEKIFYVFLPTCTNAHMGVGDFQHVQTLAYNGAFSHHFTREDDI
LDNLENGTLRLEILSDLRPTVGMIRDLSAFMTCPTTTRAVRVSDNDVTQQLAEP
ADKRTQTVLVNGVIAFAFSEFTAVTQCLEHAIPHEMYGDPRAVATMHQDVAFTVM
RNPQRAVEANRPEQLFAEYREHRSMPGKIAECLPSLSISGMTAMHKMSPAY
IAQAQKILHPGAMVTVRTDEILISENLIFFSSRATSMFTGTPNRSREARVDAVTEV
HHEWASIDTGLSYSTMTPARVAIITDMGITHQDEFFSFPFAEFGNQVNDYIKAVY
CAQRNCTLLRPRTYIAGTNVNGVGLCHGQQRTCEIIVTPVTADVAIFQKNSPRG
RAQVGSYCENTYNAEGLIYDHSRDPDAEYRSVWNPASGLSGDITMNSSTYQT
AVPGLSPCRAFTKEELIRNNGLYNNWEYSQRLGGHPATSNTEQVFVIAGTVDF
LEQCSFLOEAFALASASRALIDFNFSVKQTHAP IYGHYIIEVAVPVRRILKFNK
VF"
complement(133332)
misc_feature
/notes="DONOR: AAGGTGGTT possible donor"
complement(133352)
promoter
/notes="TATA: TATTAAA before BDLF1"
complement(133386)
promoter
/notes="TATA: TATATAA"
misc_feature
135178

```

```

promoter
/notes="polyA signal: AATAAA"
135394
/notes="TATA: TATAAGT"
misc_feature
136624
/notes="polyA signal: AATAAA"
misc_feature
136868
/notes="BAM: Bam HI D/c"
complement(137710)
/notes="TATA: TATTAAA EHL1 promoter before BcLF1, gives
4.5kb late RNA"
137857
/notes="TATA: CATAAAC"
137862..139715
/notes="BcRF1 reading frame"
138019
/notes="BAM: Bam HI c/b"
139352
/notes="BAM: Bam HI b/T"
139642..140916
/notes="BcRF1 reading frame. Northern blots detect 0.95
late and 3.8kb early RNA"
complement(140902)
/notes="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
complement(140916..143036)
/notes="BcLF2 late reading frame, encodes gp85; homologous
to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV);
NCBI gi: 833016"
/codon_start=1
/translation="MQLLCVCLVLLMEVGAASLSEVKLHLDIEGHASHYTIPTWELM
AKVCPGLEALMREANVTEDLASMNRKLIYKTSGLTGIALAEPVDIPAVSGSNQV
DASKVHPGVISGLNSPACMLSPALEKQLFYITGTMLPNTPHSPYVYQLRCHLSYVAL
SINGDKQYTGAMTSKLMGTYKRYTEKDEHVLISLVFGKTKDLPDLRGPSPYSLTS
AASGDSYLVITVTVHVFANFVFNPKDMESRAVMTFAASYAVYLQKLVILLEKMG
GCPELDTETLTTFVSVAFVFGVAGVETGCGVDLRLWAKSFFELTVTKDIIIGI
CYGATVGMOSYGLERLAAMLAATVKEELGHTTEKQYALRALATVGYKAGVYSGL
IGCATSVLSAYNRHPFQPLHTVWRETIFIGSHVVIRELRNLNVTCQPNALYQLLS
TALCALEIGEIVGLAIGTSGFSPCYLSLRFDLTRDKLMSMAPQATLDQAASN
AVDGF LARLSLEREDRAHLPAYKVDRLDKVLMIP LINVTFIISDRBVRGSALY
EASTYLLSSLSLSPVMKCSQCAVAGEPRQIPKIQNFTRTKSCIFCGFALLSDE
KGLLETTYITSQEVQNSILSNFYEDNLVHLYLLLTGTGTMEIAGLYEERAHVWL
AIIYFAFALGIFLVHKIVMFFL"
140970
/notes="polyA signal: AATAAA"
complement(141286)
/notes="polyA signal: AATAAA"
142589
/notes="TATA: CATAAAA"
142740
/notes="BAM: Bam HI T/X"
complement(143038..144861)
/notes="BcLF1 early reading frame, thymidine kinase
(Littler et al, 1986). Weak homology to RF 36 VZV and HSV
thymidine kinase. 4.0kb early RNA presumably encodes the
TK. Also a 2.2kb late RNA here. NCBI gi: 833018"
/codon_start=1
/translation="MAGFPCKEAGPPCGWRKKQCEDESPENRHNENFYAIEDDFAPSVL
TPTGSDSGAGEDDDGLYQVPTWPPIMAPTGLSGERVPCRTQAATVNTNGSPSRH
TSCPTLPRCAQPPAPAHQKPTAPTPKPSRCEGPKTDPDFSWFRKTSCTEGGAUST
SISFYKQKGEFEGLAGLDDKSDCESEDSNFRPPSSHSAIKQKNGCKGPKSGFSEH
HVPVTKPACSLFELCAPGVGKTMNLNHLKAVFGDLTIWPEPMRYWTHVYENAIKAWH
KNVTARHGREDTSAEVLACQMKFTTTFRVLASRKRSLLVETSGARSVAPLDCHLHLD
RHLLSASVPEFLMLIRSQLSLSYDFIQVLAFTADPGDITVMMKLNVEENRRKRRG
RKHESGLDGLYKSNDAHYAVYCAWLLITQYFAPEDIVKVCAGLTIITTVCHQSHPTI

```

```
IRSVAEKUYKNSIFSUIKEVYIQFPRADAVLLEVCIAFTRTTLAYLQFVLVLDSEFQDD
LPGCWEIYIMQALKNPAIRSOFFDMAGLSKVISDFERNDR"
complement (143310)
/notes="TATA: TATAAGA ECL2 late promoter before BXLF2,
143608
gives 2.5kb late RNA"
/notes="polyA signal: AATAAA"
144791
/notes="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before
BXRF1"
144860..145606
/notes="BXRF1 late reading frame, homologous to RF 35 VZV.
Basic (core?) protein. NCBI gi: 833017"
/codon_start=1
/translation="MDPTRGLICALSTHDLAKFHS LPPARKAAGKRAHLRCYSKILLSIK
SMEQLASFSLPPGPTFTDFRLFFVTLGRR TADCVVVALQF YPRCY IVEFKTAMST
ANQSVTRKAQRLEGTALCUCANFLRTSCPPVLGSGGLEVLALVFNQKRSLETLQV
EFLALGQKTLPTSTTGLIMLJLSRWQDGA LARLDLRPRPTAQGHRPRTHVGRKPSQLTA
RVPRSRAGRAGGRKKQGVAVGQVCPQAK"
144862
/notes="BAM: Bam H1 X/V"
144945
/notes="DONOR: CAGGTAAGC possible donor at 3' BXRF1"
complement (145135)
/notes="TATA: TATAACA before BXLF1"
145302
/notes="TATA: TATTTAA before BVRF1, potential promoter for
1.9kb early RNA"
145416..147128
/notes="BVRF1 early reading frame, homologous to RF 34 VZV;
NCBI gi: 833019"
/codon_start=1
/translation="MALSCHVLDIPARLPRTGCELMWAPSIRNSLRVSPFALEAER
EAEKRSERKORCAQVKNRRLVRDELGIMRDHIAAREETIRODLDAVAFSDGLSNQ
VRSFSTGSRAPAPSPSPAPQFRLTGNQAVYSISPTDPP LMAVAGSLAQTLGNIY
GNIOWVPSFGPWVRTMSANAMORRVPRKLRGNLNFNTSVSLKLMTEVAVLGGTTQ
DFFSDVRHLDPDLQALLLSVAYLLQGGSSHQRRPLPASREELLELPESLEKIIADL
RAKSPGNGRDLITSGNKEARQSIAPLNROAYPEGTADNKTYNLFVCGACLIPTAAL
NVPCAGRDRLIVRIANOIFGEDVPVPSHOMNLRVGLAALEALMIVYTLCEATIAL
EAAERLHLSLLQAMQRRKPPAMASAGMPGAYPQTLFRHGELFRE IWAHVVRPTVA
ADPQASISLFPGLVLALEIKMDQNP SHYAINLTGQKQDTLFEIINQKILFHPDA
AMLAARTQLRAFEDGVGVALGRP SPMLAAREILLERQFSASDDYDRUYFLTUGYLASP
VAP"
complement (146926)
/notes="polyA signal: AATAAA"
147167
/notes="DONOR: AAGGTAAAT possible donor"
147170
/notes="polyA signal: AATAAA, 3' end of 2.4kb late and
1.9kb early RNAs"
147721
/notes="TATA: TATTAT before BVRF2, potential promoter for
2.1kb early RNA"
147927..149744
/notes="BVRF2 early reading frame, N-terminus homologous to
RF 33 VZV; NCBI gi: 833020"
/codon_start=1
/translation="MWQAPSVVCGFVERPDAPPKQACLAHLDP LTVKSQLP LKKPIPL
TVEHLDPADPVGFLYQSRAGLESAA SITSCDFLSLIDSIYHDCDIAQSRLPIRE
PKVEALHAWLSLSLASLHPDIPQTTADGKLSFFDHVSICALGRRGTAVYGTGIDIA
WLKHFSLELPSIAAQIENDANAAKRSGCPEDHPIPL LTKLIKAIADAGFLNRNVEFL
RQDRGVANIAPAESYLKADAPDLQKPKALQPPASTDPATMLSGNAGEGATACGGS
ANAQQLISVPRNFTETLIQTNLDNKP PQTP LPYAAP LPFPPSHQAIATAPSYGAG
AVAPAGGYFTSPGCVYAGPAGGDP GAF LAMDAHTYHPHPPPAYFGLGFLGPPPPV
```

```
PPYSGHRLADYVPAPSRNSKRKRDRPEDEEGGLFPGEDATLYRKDIAGLSKSVNEL
QHTLAARETISYHTGVGCPQCPCTYHSQFQFQHSYEVRYVPVPPPPPTS
HOAAQAPPPPCPTQAPAEAHCAVESTIPEAGAACNSCPREDTNPQPTTEGHHHKKLV
QASASGVAQSKEPTTPKAKSVSAHLKSI FCEELLNKRVA"
148007
/notes="BAM: Bam H1 V/d"
148620
/notes="TATA: TATTTAA ECR1 late promoter before BdrF1,
gives 1.2kb late RNA"
148707..149744
/notes="BdrF1 reading frame; this is the C terminus of
BVRF2; NCBI gi: 833021"
/codon_start=1
/translation="MLSGNAGEATACGSAAGQQLISVPRNTFTLLQTNLDNKP RP
RQTP LTYAAP LPFPF SHQAIATAPSYGAGAVAPAGGYFTSPGCVYAGPAGGDPGAF
AMDATYHPHPPPAYFGLGFPVPVPPYVYSGHRLADYVPAPSRNSKRKRDPPEE
DEEGGLFPGEDATLYRKDIAGLSKSVNELQHTLAARETISYHTGVGCPQCPCC
YTHSGPYGFPQHSYEVRYVPVPPPPTS HQAAQAPPPPTQAPAEAHCAVESTIPE
AGAACNSCPREDTNPQPTTEGHHHKKLVQASASGVAQSKEPTTPKAKSVSAHLKSI
FCEELLNKRVA"
149115
/notes="BAM: Bam H1 d/I"
149727
/notes="polyA signal: AATAAA, 3' end of 2.1kb early and
1.2kb late RNAs"
complement (149758)
/notes="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb
late and 1.8kb late RNAs"
complement (149779..150525)
/notes="BILF2 late reading frame 11XNTT/S; NCBI gi:
833022"
/codon_start=1
/translation="WTHLVLLCCCVSGCAFFSDLVKEENVTAHAGARNVLTCSVPS
NEVSRIELGRGTYTQDQLPLAVATSNNGHTITNGGNYSLTLEWVNDSTNSLLI
PNVTLAHGYTQNTVLNGSVASGVHGNYSAGEDDQYHANRTLTORMHLTVIPATT
IAPTTLSHTSTSHRPHRPVSKRPTUKQVTLGFPF IDPWRPKTTWVHALLLLITCA
VAPVLLIIIIISCLGMLAGRRRKGHPL"
complement (150571)
/notes="TATA: TATTTAG before BILF2. Potential promoter for
1.0kb late RNA."
151236..151618
/notes="repetitive sequence 3X"
151767
/notes="polyA signal: AATAAA"
complement (151780)
/notes="TATA: CATAAAA"
152012..152013
/notes="DEL: B95-8 deletion with respect to Raji"
complement (152161..153099)
/notes="BILF1 reading frame, membrane protein, 3xNXS/T;
NCBI gi: 833023"
/codon_start=1
/translation="MLSTMAPGSTVGTUVANWTSVNATEDACTKS YAFISGWTSLLL
VLLIILLTAGLIFTIFVRKLVHRMDVWLIALLIELLLMVLGRMIQFSTSGCLLTQN
MFLGLMCSVWTHLGMALXETIALFSRTPKRTSHRNVCLYLMGVFCVILLIILIT
MGEDANLGNPNRCREGTCKMGTAVQCIKAGCYLLAAVLIVLTVIIWKLLRTKEG
RKRPLIGNVTFTGLICAFSWMFLSLP LILFGEAGSLGFDCTESLVARYYPPGACIAL
LIIILYAMSFSEHMDSLKNQVTVTARYPRVPVQSQT"
152230
/notes="TATA: CATAAAA"
153259
/notes="polyA signal: AATAAA"
153637
/notes="HPN: 22bp 2-fold symmetric"
```

```

misc_feature
/notes="DONOR: AAAAGTCAGC possible donor"
complement(153690)
CDS
/notes="BALF5 DNA polymerase (early), homologous to many
DNA polymerases, CMV HFLF2 and RF 28 VZV. 4.5kb early RNA
apparently encodes BALF5, RNA ends unknown. NCBI gi:
833024"
/codon_start=1
/translation="MSGGLFYNPPLRPNKGLLRKPKEYLRLIPKCFQTPGAAGVDV
RGPQPLCFYODSLTVVGGDECKGMWRQRAQGTAREADTHGSPIDFHVYDILET
VYTHEKAVIPESKQGVVPCGIVIKLLGRKADGASVCWVFGQAIFYASAPQGLD
WEFASALKASTIEDRRPTCRVSEKVTTRSGIMYGNHAGDYHKITLISHPSNVCHAT
VLADKGRICFEANVDATRFVINDVFTFGWYSCRRAPRLQHRDSYAEIYDCVCG
DLVSREDSMP$QALAFDIECLGEGEPPATNEADLLIQISCVLWSTGEEAGRYR
LILTLGTCEDETEGVEYEPSELDMLYAFFQLIRDSLVEIVTGVNANFDMPTLIDRA
RHYSINPASKGIRAGGVCVRPHDCKGFLRANTKVRITGLIPIDMYAVCRDKLS
LSDYKLDTVARHLIGAKEDVHYKEIPRLFAAGPEGRRILGMVCQDSALWMDLHIF
VIHVEAEAKIAHPIPCRVLDGQQRVFSCLLAAAKENFTILPMP$ASRDGQYGA
TVIQLSGFYNSPVLVDVFPASLYPSIIQAHNLCTSTMTPEGEHRLAGLRPEDYESF
RLTGGYVHFVKVHVESFLASILTSWLAKRKAIKKLLAACEDPQRQTILDQQLAKC
TCNAVYGTGVANGLEPCLSIAETVTLQGTTRMLERAKAFVEALSIPANUALAPSPAW
APLNPEQGLRITYGDTSDFTECRGFSESETLRFADALAAHTSLFVAPISLEAKT
FSCMLITKRYVGVLTIDCKTLMKGVELVRKTACKVFQTRCRRVLDVLADARVKEA
SLASHRPFQESFGQLPGVLPVITIMQAVTIDLRGVPMGELCFSTELSRKLSATK
STQMPHAPYQKORFVERNEELPQIHDIRIQYVFVEFGKVGKARKTEMAEDPAYAERGV
PVAVDHYFDKLQGAANIILQCLFDNNSGAALSVLQNFATRPFF"
154747
misc_feature
/notes="BAM: Bam H1 I/A"
complement(156707)
misc_feature
/notes="polyA signal: AATAAA; 3' end of 2.5kb late (gb) RNA
and 1.8kb late RNA"
CDS
complement(156749..159322)
/notes="BALF4 late reading frame 9xNXT/S homologous to HSV1
glycoprotein B (Pellet et al, 1985), CMV HFLF1 and RF 31
VZV (gpII); NCBI gi: 833025"
/codon_start=1
/translation="MTRRVLVVVLLAALACRLGAQTPEQPAPPAATTVPQTATROQT
SFFPRVCELSHGDLFESSDIQCFSGTRENHTEGLMVFKNIIIPYSFKRVSYTKI
VTNLIYNGVADSVTRNEEKE$YDSYETQMDTIVQCYNAVKMTKGLTRVYVDRI
GVNITVNLKPTGGLANGVRVYASQTELDAPGHLINTYRTYTNCLITDMAKSN$P
FDFEFTTTCQTVKESPFYDGKQKETFHERADSFHRTYTKIVDYDRGNTNPQGERAF
LDKCTVTLSWKLNRHTAYCPLQHMOTFDSTIATETGKSIFHVTDEGTSFVNTTVGI
ELDPAKCIEQVQNK$MHEKYEAVDQRYTKQEAITYFTISGGLLAWLPITPRSLAT
VKNLTELTPPTSSPSPSPAP$AARGSTPAAVIRRRRRDAGNATTPVPTAPGKSL
GLNNPATVQIQFAYDSLRAQINRMGLDLARAWCLEQRQNMVRELTKINPTWMS
IYGVAAKRLGDVIVSQCVPVQATVTLRKSNRVPCSE$WICYSRPLV$FSFINDTK
TYEQGLGTNEIFLTKKTEVCQAT$QYFQSGNEIHVNDYHHFRTIELDGIATLQT
FISLNTLSTENIDFASLEIYSRQDRASNVDFLEGIFREYNFQAOINAGRLQDLN$V
SNGBNQFVDGLGELMDSLGSVCQSI$TNLVSIVTGGLESSIV$SCFISFFKNPFCGMLIIV
LVAGVILVILSTRTRQM$SQPVQMLPGIDELAQHASGEGPGINPISKTELQAIM
LALHEQNO$QKRAAQAAGP$VASRALQAARDREPGLRRRRYHDPETAAALLGEAETE
F"
158204
promoter
/notes="TATA: TATAAAA"
CDS
complement(159312..161678)
/notes="BALF3 reading frame"
promoter
complement(159370)
/notes="TATA: TATTTAA ECL1 late promoter before BALF4,
gives 2.5kb late RNA"
160966
misc_feature
/notes="polyA signal: AATAAA"
160990
misc_feature
/notes="polyA signal: AATAAA"
misc_feature
complement(161013)

```

```

CDS
/notes="polyA signal: AATAAA, presumed end of 3.9kb early
RNA"
complement(161384..164770)
/notes="BALF2 early reading frame, homologous to RF 29 VZV
and major DNA binding protein HSV. 3.9kb early RNA; NCBI
gi: 833026"
/codon_start=1
/translation="MQGAQTSEDNLG$SQSGPGCGYIIFYPLATYPLREVATIGCTYA
GHRCTVPLILCGITVEGFSINVKALHRRDPNCGLLRATSYHRDITYVFNHMHMPPPI
FEGPGLGALCGTREVTCYDAYSALPRESKPGDFEPGLDPSAYLGAVAITAEKFER
LYSGLNVAI$PSLKAEVAVQ$SASVRPLYDKVEFPEGVQ$LRQFYNSDLSRCH$EALY
TGLAQALRRVRKGLKELLEK$SQ$QAKVAKVAPLKEFPASTISHPDSGALMIV$D$A
ACELAV$YAP$M$EAS$HET$PASIN$Y$W$P$LEADCE$PEARV$AALHRYNASI$APH$V$Q
IFATNSVLYVSGVSK$SQ$Q$K$ESL$N$SY$M$H$G$LT$Q$E$G$T$W$D$P$C$R$C$F$S$M$G$G$D$V
TCTNGP$N$Y$A$V$E$H$V$Y$A$S$F$N$L$R$A$Y$Y$V$Y$Q$Q$Q$S$L$T$P$V$E$T$S$Y$V$A$A$A$S
P$M$C$S$C$E$G$R$A$P$A$V$C$L$N$T$F$R$L$D$R$E$P$V$M$S$T$O$R$D$P$Y$V$I$S$G$S$Y$N$E$T$D$F$L$G$N$F$A$N
F$D$K$E$D$Q$Q$R$P$D$E$P$R$Y$T$W$Q$N$L$L$E$R$L$S$R$L$G$I$D$A$E$G$K$E$K$E$P$H$G$P$R$D$F$V$M$K$F$D$V
D$A$A$V$D$E$V$Q$F$N$S$M$A$K$N$N$I$T$Y$K$D$L$V$K$S$Y$H$V$M$Q$Y$S$C$N$P$A$Q$P$A$C$P$I$F$T$Q$L$F$Y$R$S$L$T
I$Q$D$S$I$S$P$I$C$M$C$Y$E$N$D$N$P$G$Q$S$P$P$E$M$K$G$H$Y$Q$T$L$C$T$N$F$R$S$L$A$D$K$G$V$L$T$A$K$E$A$K$V$H
G$E$P$T$C$D$P$D$D$A$A$Q$R$V$Y$G$R$L$P$V$M$S$K$V$M$A$L$C$P$N$K$I$K$N$R$V$V$T$C$E$N$A$A$L$N$S$F$I
K$S$T$T$R$E$N$Y$I$N$G$P$Y$M$K$F$A$T$Y$H$K$T$L$F$P$D$T$K$S$S$Y$L$H$M$F$S$R$R$S$V$P$V$P$S$G$A$S$E$Y
S$D$A$L$F$V$D$G$S$R$A$H$E$S$N$V$D$V$P$N$V$Y$T$V$A$K$Q$R$L$N$N$A$I$L$K$A$C$Q$Q$T$Q$F$Y$S$I$Q$G$L$V$P
R$Q$S$P$A$R$D$Y$P$H$V$L$G$T$R$A$V$E$S$A$A$Y$A$E$A$T$S$S$L$T$A$T$T$V$C$A$N$D$C$L$S$Q$V$C$K$A$R$V$W$T$P$
V$T$N$K$Y$T$G$V$G$N$N$Q$IF$Q$A$N$G$L$F$M$G$R$V$D$N$D$L$Q$A$C$A$G$L$R$Q$A$G$G$S$M$R$K$K$F$V$F$A$T
P$T$G$I$T$V$K$R$T$Q$A$A$T$T$E$N$R$A$G$L$E$A$I$S$Q$K$E$E$D$C$V$D$V$N$C$V$N$L$V$D$M$C$E$A$S$I$T
R$D$A$E$Y$L$G$R$F$S$V$L$A$D$S$V$L$E$T$A$T$A$S$S$G$I$E$W$T$A$E$A$R$D$F$L$E$G$W$G$P$G$A$Q$D$N$F$I$S$V
A$E$P$V$T$A$S$Q$A$S$A$G$L$L$G$G$G$G$G$G$R$R$R$R$A$T$V$I$P$G$L$E$V"
163978..166635
/notes="DEL: deletion in Raji"
complement(164814)
promoter
/notes="TATA: CATTTAA before BALF2, presumed promoter for
3.9kb early RNA"
misc_feature
complement(164851)
CDS
/notes="polyA signal: AATAAA"
complement(164855..165517)
/notes="BALF1 early reading frame, 0.7kb early RNA; NCBI
gi: 833028"
/codon_start=1
/translation="MNLAIALDSRHPGGLASYTILPRPFYHISILKPVSWPDETMRPAKS
TUSVVRTPVEAWAPSPDDKVAESSYLMFRAMVAVFTRDEKDLPLPALVLCRLIKA
SLRKORKL$YAE$ACR$TAD$IG$G$D$H$V$R$L$I$T$S$V$L$R$A$V$N$D$H$Y$W$S$R$L$R$V$W$I$C$T$V$W$F$A
V$N$Y$D$D$H$K$A$A$F$V$L$G$A$T$A$H$Y$L$A$Y$R$R$L$W$F$A$R$L$G$M$P$R$S$L$R$Q$F$P$V$T$W$A$S$I$D$F$I$K
S$"
165442
promoter
/notes="TATA: GATAAAA"
165466
promoter
/notes="TATA: TATAAGA EDRI early promoter before BALF1,
gives 0.8kb early RNA"
165504..166169
CDS
/notes="BALF1 reading frame; NCBI gi: 833027"
/codon_start=1
/translation="MARFIAQLILLASCVAAAGQAVTAFIGERVTLTSYVRRVSLGPEI
EVSWEKLGPEEQVLI$G$M$H$D$V$I$E$N$P$F$R$G$F$D$H$R$S$A$N$T$F$F$L$V$T$A$N$A$I$S$H$D$N$Y
L$C$R$M$K$G$E$T$Q$K$E$H$S$V$W$K$P$L$I$T$S$H$S$E$R$S$Q$F$D$F$S$V$T$V$T$C$T$V$N$A$F$P$H$P$H$Q$M$IM
P$E$G$V$E$P$A$P$T$A$A$G$G$V$M$K$E$D$G$S$V$A$V$D$S$L$P$K$P$H$L$P$V$T$C$V$G$K$N$D$K$E$A$H$G$V$Y$V$S$G$Y
L$Q$"
complement(165713)
promoter
/notes="TATA: TATAAAG before BALF1"
166165
misc_feature
/notes="polyA signal: AATAAA 3' end of 0.8kb early RNA.
Also 1kb late RNA in this region."
166469..166475
promoter
/notes="TATA: TTATTTT"
166498..166916
mRNA

```


sapiens (clone library: Kai Wang's) (tissue library: ATCC 1521)
DNA; Homo sapiens DNA; Homo sapiens DNA; Homo sapiens
sapiens (clone library: Kai Wang's) DNA; Homo sapiens (clone
library: Kai Wang's) DNA; Homo sapiens (clone library: YAC 234
A6f6) DNA; Homo sapiens (clone library: YAC 234 A6f6) DNA; Homo
sapiens (clone library: YAC 234 A6f6) DNA; Homo sapiens DNA; Homo
sapiens DNA; Homo sapiens DNA; Homo sapiens (clone library: YAC 234
A6f6) DNA; Homo sapiens (clone library: YAC 234 A7B23) DNA; Homo
sapiens (clone library: Kai Wang's) DNA; Homo sapiens (clone
library: YAC 199 B157C9) DNA; Homo sapiens (clone library: YAC
A7B23) DNA; Homo sapiens (clone library: Kai Wang's) DNA; Homo
sapiens (clone library: Kai Wang's) (tissue library: ATCC 1521)
DNA; Homo sapiens (clone library: Kai Wang's) DNA; Homo sapiens
DNA; and Homo sapiens (clone library: Kai Wang's) DNA.

Homo sapiens

Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Homnidae.

1 (bases 1 to 684973)

Rowen, L., Koop, B.F. and Hood, L.
Sequence of the human T cell receptor beta locus
Unpublished (1994)

2 (bases 368065 to 442094)

Slightom, J.L., Siemieniak, D.R., Siu, L.C., Koop, B.F. and Hood, L.
Nucleotide sequence analysis of 77.7 kb of the human V beta T-cell
receptor gene locus: direct primer-walking using cosmid template
DNAs

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

repeat_unit	/note="putative" /rpt_family="LINE 1" (35482..353482)..353798 /note="putative" /rpt_family="MSTA" 36258..72769 /notes="(vector PWE15A)" /organism="Homo sapiens" /cell_line="CGM1" /germline /sequenced_mol="DNA" /clone="K35" /clone_lib="YAC D49H4" /citation=[1] /notes="putative"
source	/rpt_family="Alu" /notes="(vector PWE15A)" /organism="Homo sapiens" /cell_line="CGM1" /germline /sequenced_mol="DNA" /clone="K26" /clone_lib="YAC D49H4" /citation=[1]
repeat_unit	/note="putative" /rpt_family="Alu" complement(41973..42235) /notes="putative"
repeat_unit	/rpt_family="Alu" complement(42818..43079) /notes="putative" /rpt_family="Alu" 45446..45721 /notes="putative"
repeat_unit	/rpt_family="Alu" complement(49022..49282) /notes="putative" /rpt_family="Alu" complement(51129..51357) /notes="putative" /rpt_family="MLT1A" complement(52099..54780) /notes="putative"
repeat_unit	/rpt_family="LINE 1" 54870..54937 /notes="putative" /rpt_family="MER25" complement(56562..56788) /notes="putative"
repeat_unit	/rpt_family="Alu" 58592..58862 /notes="putative"
repeat_unit	/rpt_family="Alu" 60007..60287 /notes="putative"
repeat_unit	/rpt_family="Alu" 67296..71811 /notes="putative"
repeat_unit	/rpt_family="LINE 1" 70574..70940 /notes="putative"
source	/rpt_family="THE transposon like element" 71776..107249 /notes="(vector PWE15A)" /organism="Homo sapiens" /cell_line="CGM1" /germline /sequenced_mol="DNA" /clone="K26" /clone_lib="YAC D49H4" /citation=[1] complement(74111..74372) /notes="putative" /rpt_family="Alu" 77536..77883 /notes="putative"
repeat_unit	


```

repeat_unit      /rpt_family="MLTIA"
                  complement(78720..78991)
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  complement(79236..79567)
                  /note="putative"
mRNA             /rpt_family="MSTA"
                  complement(join(<79799..79948,80489..80625,81147..81400,
82186..82348,84121..84159))
                  /gene="TRYF"
exon             /gene="TRYF"
                  complement(<79799..79948)
                  /gene="TRYF"
                  /note="putative"
                  /pseudo
                  /number=5
CDS              complement(join(79799..79948,80489..80625,81147..81400,
82186..82348,84121..84159))
                  /gene="TRYF"
                  /note="putative"
                  /pseudo
intron           /codon_start=1
                  complement(79949..80488)
                  /gene="TRYF"
                  /note="putative"
                  /number=4
exon             complement(80489..80625)
                  /gene="TRYF"
                  /note="putative"
                  /pseudo
intron           /number=4
                  complement(80626..81146)
                  /gene="TRYF"
                  /note="putative; does not fit consensus"
                  /number=3
exon             complement(81147..81400)
                  /gene="TRYF"
                  /note="putative"
                  /pseudo
                  /number=3
intron           complement(81401..82185)
                  /gene="TRYF"
                  /note="putative"
                  /number=2
exon             complement(82186..82348)
                  /gene="TRYF"
                  /note="putative"
                  /pseudo
                  /number=2
intron           complement(82349..84120)
                  /gene="TRYF"
                  /note="putative"
                  /number=1
repeat_unit      82873..83151
                  /note="putative"
                  /rpt_family="Alu"
exon             complement(84121..84159)
                  /gene="TRYF"
                  /note="putative"
                  /pseudo
                  /number=1
repeat_unit      complement(85009..85072)
                  /note="putative"
                  /rpt_family="MIR"
repeat_unit      85538..85773

```

```

repeat_unit      /note="putative"
                  /rpt_family="Alu"
                  complement(86349..86618)
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  complement(90001..90268)
                  /note="putative"
V_segment        /rpt_family="Alu"
                  join(91557..91597,91715..92006)
                  /partial
                  /gene="TCRBV27S1"
                  /note="putative"
                  /pseudo
                  /codon_start=1
                  <91557..91597
exon             /gene="TCRBV27S1"
                  /note="putative"
                  /pseudo
                  /number=1
intron           91598..91714
                  /gene="TCRBV27S1"
                  /note="putative"
                  /number=1
exon             91715..92006
                  /gene="TCRBV27S1"
                  /note="putative"
                  /pseudo
                  /number=2
misc_signal      92007..92013
                  /gene="TCRBV27S1"
                  /note="RSS heptamer - awaiting approval of new feature
key; putative"
misc_signal      92007..92045
                  /gene="TCRBV27S1"
                  /note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal      92014..92036
                  /gene="TCRBV27S1"
                  /note="RSS spacer - awaiting approval of new feature key;
putative"
                  92037..92045
misc_signal      /gene="TCRBV27S1"
                  /note="RSS nonamer - awaiting approval of new feature key;
putative"
                  complement(92611..92867)
                  /note="putative"
                  /rpt_family="Alu"
                  93469..93478
misc_feature      /gene="TCRBV22S1"
                  /note="conserved decamer"
                  join(93552..93600,93689..93986)
V_segment        /gene="TCRBV22S1"
                  /note="putative"
                  <93552..93600
exon             /gene="TCRBV22S1"
                  /note="putative"
                  /number=1
intron           93601..93688
                  /gene="TCRBV22S1"
                  /note="putative"
                  /number=1
exon             93689..93986
                  /gene="TCRBV22S1"
                  /note="putative"

```

```

source
/number=2
93930..130694
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="K56"
/clone_lib="YAC D49H4"
/citation=[1]
93987..94025
/genes="TCRBV22S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
93987..93993
/notes="RSS_heptamer - awaiting approval of new feature
key"
93994..94016
/notes="RSS spacer - awaiting approval of new feature key"
94017..94025
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(95946..96207)
/notes="putative"
/rpt_family="Alu"
complement(98840..99104)
/notes="putative"
/rpt_family="Alu"
99921..99960
/notes="polymorphic"
/rpt_family="microsatellite, gt"
/rpt_unit=99921..99922
/evidence=experimental
101036..101045
/genes="TCRBV9S1"
/notes="conserved decamer; putative"
join(101150..101198,101315..101609)
/genes="TCRBV9S1"
/notes="putative"
<101150..101198
/genes="TCRBV9S1"
/notes="putative"
/number=1
101199..101314
/genes="TCRBV9S1"
/notes="putative"
/number=1
101315..101609
/genes="TCRBV9S1"
/notes="putative"
/number=2
/codon_start=3
101610..101616
/notes="RSS_heptamer - awaiting approval of new feature
key"
101610..101648
/genes="TCRBV9S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
101617..101639
/notes="RSS spacer - awaiting approval of new feature key"
101640..101648
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(104481..104664)
/notes="putative"

```

```

misc_feature
/rpt_family="Alu"
105669..105678
/genes="TCRBV7S1"
/notes="conserved decamer; putative"
join(105774..105822,105933..106227)
/genes="TCRBV7S1"
/notes="putative"
<105774..105822
/genes="TCRBV7S1"
/notes="putative"
/number=1
105823..105932
/genes="TCRBV7S1"
/notes="putative"
/number=1
105933..106227
/genes="TCRBV7S1"
/notes="putative"
/number=2
106228..106266
/genes="TCRBV7S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
106228..106234
/notes="RSS_heptamer - awaiting approval of new feature
key"
106235..106257
/notes="RSS spacer - awaiting approval of new feature key"
106258..106266
/notes="RSS_nonamer - awaiting approval of new feature key"
106518..106791
/notes="putative"
/rpt_family="Alu"
complement(107940..108191)
/notes="putative"
/rpt_family="LINE 1"
complement(110274..110535)
/notes="putative"
/rpt_family="Alu"
113509..113518
/genes="TCRBV5S1"
/notes="conserved decamer; putative"
join(113622..113670,113798..114091)
/genes="TCRBV5S1"
/notes="putative"
<113622..113670
/genes="TCRBV5S1"
/notes="putative"
/number=1
113671..113797
/genes="TCRBV5S1"
/notes="putative"
/number=1
113798..114091
/genes="TCRBV5S1"
/notes="putative"
/number=2
114092..114130
/genes="TCRBV5S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
114092..114098
/notes="RSS_heptamer - awaiting approval of new feature
key"

```



```

misc_signal 138532..138538
/notes="RSS heptamer - awaiting approval of new feature
key"
misc_signal 138532..138570
/genes="TCRBV7S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
138539..138561
/notes="RSS spacer - awaiting approval of new feature key"
138562..138570
/notes="RSS nonamer - awaiting approval of new feature key"
replace(138563,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(138639,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139298,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139322,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139650,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139980,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140057..140058,"at")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140126..140129,"ta")
/notes="found in cosmid clone HVB15 in L36190"
replace(140236,"t")
/notes="found in cosmid clone HVB15 in L36190"
replace(140271,"t")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140374,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140399,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
(140495..140709)..(162290..162390)
/notes="hypothetical location of insertion/deletion
polymorphism, breakpoints correspond to position 4141-4223
in cosmid HVB15 in accession number L36190; putative"
complement(141314..141574)
/notes="putative"
/rpt_family="Alu"
141809..141818
/genes="TCRBV13S2"
/notes="conserved decamer; putative"
join(141898..141946,142036..142330)
/genes="TCRBV13S2"
/notes="putative"
<141898..141946
/genes="TCRBV13S2"
/notes="putative"
/number=1
141947..142035
/genes="TCRBV13S2"
/notes="putative"
/number=1
142036..142330
/genes="TCRBV13S2"
/notes="putative"
/number=2
142331..142337
/notes="RSS heptamer - awaiting approval of new feature
key"
misc_signal 142331..142369

```

```

/genes="TCRBV13S2"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
142338..142360
/notes="RSS spacer - awaiting approval of new feature key"
142361..142369
/notes="RSS nonamer - awaiting approval of new feature key"
145518..145557
/notes="polymorphic"
/rpt_family="microsatellite"
/rpt_unit="145518..145519
/evidence=experimental
150631..150733
/notes="putative"
/rpt_family="Alu"
151765..151774
/genes="TCRBV9S2"
/notes="conserved decamer; putative"
<151879..151927
/genes="TCRBV9S2"
/notes="putative"
/pseudo
/number=1
join(151879..151927,152044..152338)
/genes="TCRBV9S2"
/notes="putative"
/pseudo
151928..152043
/genes="TCRBV9S2"
/notes="putative"
/number=1
151952..186019
/notes="(vector PWE15A)"
/organisms="Homo sapiens"
/germline
/sequenced_mol="DNA"
/clone="G54"
/clone_lib="Kai Wang's"
/tissue_lib="ATCC 1521"
/citation=[1]
152044..152338
/genes="TCRBV9S2"
/notes="putative"
/pseudo
/number=2
152339..152345
/notes="RSS heptamer - awaiting approval of new feature
key"
152339..152377
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
152346..152368
/notes="RSS spacer - awaiting approval of new feature key"
152369..152377
/notes="RSS nonamer - awaiting approval of new feature key"
152333..155242
/genes="TCRBV7S2"
/notes="conserved decamer; putative"
join(155338..155386,155497..155791)
/genes="TCRBV7S2"
/notes="putative"
<155338..155386
/genes="TCRBV7S2"
/notes="putative"

```

```

/number=1
155387..155496
/genes="TCRBV752"
/notes="putative"
/number=1
155497..155791
/genes="TCRBV752"
/notes="putative"
/number=2
155792..155830
/genes="TCRBV752"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
155792..155798
/notes="RSS_heptamer - awaiting approval of new feature
key"
155799..155821
/notes="RSS_spacer - awaiting approval of new feature key"
155822..155830
/notes="RSS_nonamer - awaiting approval of new feature key"
157060..157320
/notes="putative"
/rpt_family="Alu"
replace(162571..162572,"cc")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162574,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162601,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162622,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162763,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162969,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
complement(162995..163257)
/notes="putative"
/rpt_family="Alu"
replace(163006,"t")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163460,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163479,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
163490..163499
/genes="TCRBV13S9"
/notes="conserved decamer; putative"
replace(163552,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163562,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
join(163579..163627,163717..164011)
/genes="TCRBV13S9"
/notes="putative"
<163579..163627
/genes="TCRBV13S9"
/notes="putative"
/number=1
replace(163604..163607,"cagc")
/genes="TCRBV13S9"
/notes="putative"
163628..163716
/genes="TCRBV13S9"
/notes="putative"

```

```

variation
/number=1
replace(163641,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163652..163653,"ca")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163669..163671,"ag")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163677,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163680..163681,"gt")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163685,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163690,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163705,"t")
/notes="found in cosmid clone HVB15 in L36190; putative"
163717..164011
/genes="TCRBV13S9"
/notes="putative"
/number=2
164012..164018
/notes="RSS_heptamer - awaiting approval of new feature
key"
164012..164050
/genes="TCRBV13S9"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
164019..164041
/notes="RSS_spacer - awaiting approval of new feature key"
164042..164050
/notes="RSS_nonamer - awaiting approval of new feature key"
replace(164142,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
164352..203210
/notes="(vector PTL5)"
/organism="Homo sapiens"
/cell_line="HeLa"
/germline
/sequenced_mol="DNA"
/clone="H137"
replace(164472,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(164630,"c")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(164983,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165007,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165593,"c")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165698,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165810,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165949,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in

```

```

variation      L36190; putative"
                replace(165972,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(165986,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166193..166195,"ta")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166294,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166310,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166389,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166417..166421,"ccaat")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166428,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166433,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166501,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166508,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166529,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166896,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166913,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                167102..167111
                /gene="TCRBV6S7"
                /note="conserved decamer; putative"
                replace(167136,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
V_segment      L36190; putative"
                Join(167212..167260,167411..167708)
                /gene="TCRBV6S7"
                /note="putative"
                <167212..167260
                /gene="TCRBV6S7"
                /note="putative"
                /number=1
                replace(167260,"t")
                /gene="TCRBV6S7"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
allele          L36190; putative"
                167261..167410
                /gene="TCRBV6S7"
                /note="putative"
                /number=1
intron          L36190; putative"
                /note="found in cosmid H137 and in cosmid clone HVB15 in

```

```

variation      replace(167289..167306,"tt")
                /note="found in cosmid clone HVB15 in L36190; putative"
                167291..167336
                /note="polymorphic"
                /rpt_family="microsatellite"
                /rpt_unit=167291..167292
                /evidence=experimental
                167411..167708
                /gene="TCRBV6S7"
                /note="putative"
                /number=2
                replace(167532,"g")
                /gene="TCRBV6S7"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(167633,"a")
                /gene="TCRBV6S7"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                167709..167747
                /gene="TCRBV6S7"
                /note="RSS - awaiting approval of new feature key;
                heptamer is at 5' end; putative"
                167709..167715
                /note="RSS heptamer - awaiting approval of new feature
                key"
                167716..167738
                /note="RSS spacer - awaiting approval of new feature key"
                167739..167747
                /note="RSS nonamer - awaiting approval of new feature key"
                replace(167778,"g")
                /note="found in cosmid H137; putative"
                replace(168073,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168168,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168388,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168473,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168864,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169009,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169470,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169499,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169521,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169567..169569,"caaa")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169688,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in

```

```

variation      L36190; putative"
                replace(169885,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170006,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170579,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170646,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170683,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171468,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171496..171497,"ctt")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171590,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
repeat_unit    complement(171597..171873)
                /note="putative"
                /rpt_family="Alu"
variation      replace(171787,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171997,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172230,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172297,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172462,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172745,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172762,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172878,"a")
                /note="found in cosmid clone HVB15 in L36190; putative"
variation      replace(172883..172891,"ccaaagctc")
                /note="found in cosmid clone HVB15 in L36190; putative"
repeat_unit    complement(172985..173195)
                /note="polymorphic; putative"
                /rpt_family="Alu"
variation      replace(173216,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
exon           L36190; putative"
                /gene="TCRBV30S1"
                /note="putative"
                /pseudo
                /number=2

```

```

allele         replace(173390,"t")
                /gene="TCRBV30S1"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
allele         replace(173532,"a")
                /gene="TCRBV30S1"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
misc_signal    173615..173621
                /gene="TCRBV30S1"
                /note="RSS heptamer - awaiting approval of new feature
                key; heptamer is at 5' end. There is a deletion in the
                spacer and 3' end is unknown.; putative"
variation      replace(173623,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(174615..174616,"gtt")
                /note="found in cosmid H137; putative"
                complement(174943..177008)
                /note="putative"
                /rpt_family="LINE 1"
                complement(175783..176050)
                /note="putative"
                /rpt_family="Alu"
                replace(176058,"t")
                /note="found in cosmid H137; putative"
variation      replace(176452,"g")
                /note="found in cosmid H137; putative"
                replace(177403,"t")
                /note="found in cosmid H137; putative"
variation      replace(178325..178326,"gtt")
                /note="found in cosmid H137; putative"
variation      replace(178623..178626,"gtc")
                /note="found in cosmid H137; putative"
                replace(179131,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(179150,"c")
                /note="found in cosmid H137; putative"
                replace(179294..179295,"atct")
                /note="found in cosmid H137; putative"
variation      replace(179414,"g")
                /note="found in cosmid H137; putative"
                179576..179776
                /note="putative"
                /rpt_family="MER43"
                complement(180768..181166)
                /note="putative"
                /rpt_family="MSTA"
                replace(181040..181043,"ta")
                /note="found in cosmid H137; putative"
                181515..183318
                /note="putative"
                /rpt_family="THE transposon like element"
                replace(181551,"t")
                /note="found in cosmid H137; putative"
                replace(181567,"g")
                /note="found in cosmid H137; putative"
                complement(181892..182155)
                /note="putative"
                /rpt_family="Alu"
                replace(181912,"a")
                /note="found in cosmid H137; putative"
                replace(182049,"t")

```

```

variation      /note="found in cosmid H137; putative"
                replace(182640,"a")
variation      /notes="found in cosmid H137; putative"
                replace(183279,"g")
repeat_unit    /notes="found in cosmid H137; putative"
                complement(183519..183895)
                /note="putative"
                /rpt_family="MSTA"
variation      replace(183761,"t")
variation      /note="found in cosmid H137; putative"
                replace(185410,"g")
variation      /notes="found in cosmid H137; putative"
                replace(185928,"t")
variation      /note="found in cosmid H137; putative"
                187054..187063
misc_feature    /genes="TCRBV31S1"
                /note="conserved decamer; putative"
                join(187166..187214,187324..187616)
                /partial
V_segment      /genes="TCRBV31S1"
                /note="putative"
                /pseudo
                /codon_start=1
                <187166..187214
                /genes="TCRBV31S1"
                /notes="putative"
                /pseudo
                /number=1
                187215..187323
                /genes="TCRBV31S1"
                /notes="putative"
                /number=1
                187324..187616
                /genes="TCRBV31S1"
                /notes="putative"
                /pseudo
                /number=2
                187617..187623
                /genes="TCRBV31S1"
                /note="RSS heptamer - awaiting approval of new feature
                key; heptamer is at 5' end. Span of RSS not known.;"
                putative"
repeat_unit    188638..188908
                /notes="putative"
                /rpt_family="Alu"
repeat_unit    189459..189726
                /notes="putative"
                /rpt_family="LINE 1"
repeat_region  189754..189806
                /notes="polymorphic"
                /rpt_family="microsatellite"
                /rpt_unit=189754..189755
                /evidence=experimental
repeat_unit    190067..190346
                /notes="putative"
                /rpt_family="Alu"
                193517..230319
                /notes="(vector PTL5)"
source         /organism="Homo sapiens"
                /cell_type="sperm"
                /germline
                /sequenced_mol="DNA"
                /clone="H18"
misc_feature    195433..195442

```

```

V_segment      /gene="TCRBV13S5"
                /note="conserved decamer; putative"
                join(195522..195570,195662..195956)
                /gene="TCRBV13S5"
                /note="putative"
                <195522..195570
                /gene="TCRBV13S5"
                /note="putative"
                /number=1
                195571..195661
                /gene="TCRBV13S5"
                /note="putative"
                /number=1
                195662..195956
                /gene="TCRBV13S5"
                /number=2
                195957..195995
                /gene="TCRBV13S5"
                /note="RSS - awaiting approval of new feature key;
                heptamer is at 5' end; putative"
                195957..195963
                /note="RSS heptamer - awaiting approval of new feature
                key"
                195964..195986
                /note="RSS spacer - awaiting approval of new feature key"
                195987..195995
                /note="RSS nonamer - awaiting approval of new feature key"
                replace(198022,"t")
                /note="from cosmid H18; putative"
                198980..198989
                /gene="TCRBV6S1"
                /note="conserved decamer; putative"
                join(199093..199141,199252..199549)
                /gene="TCRBV6S1"
                /note="putative"
                <199093..199141
                /gene="TCRBV6S1"
                /number=1
                199142..199251
                /gene="TCRBV6S1"
                /note="putative"
                /number=1
                199173..199202
                /note="polymorphic"
                /rpt_family="microsatellite"
                /rpt_unit=199173..199174
                /evidence=experimental
                199252..199549
                /gene="TCRBV6S1"
                /note="putative"
                /number=2
                199550..199588
                /gene="TCRBV6S1"
                /note="RSS - awaiting approval of new feature key;
                heptamer is at 5' end; putative"
                199550..199556
                /note="RSS heptamer - awaiting approval of new feature
                key"
                199557..199579
                /note="RSS spacer - awaiting approval of new feature key"
                199580..199588
                /note="RSS nonamer - awaiting approval of new feature key"
                join(200895..200939,201077..201349)
                /partial
                V_segment

```



```

/genes="TCRBV32S1"
/notes="putative"
/pseudo
/codon_start=1
<200895..200939
/genes="TCRBV32S1"
/notes="putative"
/pseudo
/number=1
200940..201076
/genes="TCRBV32S1"
/notes="5' end does not fit consensus gt; putative; does
not fit consensus"
201077..201349
/genes="TCRBV32S1"
/notes="putative"
/pseudo
/number=2
201350..201394
/genes="TCRBV32S1"
/notes="RSS - awaiting approval of new feature key; spacer
sequence too long; putative"
201350..201356
/genes="TCRBV32S1"
/notes="TCRBV32S1"
/notes="RSS heptamer - awaiting approval of new feature
key; putative"
201357..201385
/genes="TCRBV32S1"
/notes="RSS_spacer - awaiting approval of new feature key;
putative"
201386..201394
/genes="TCRBV32S1"
/notes="RSS_nonamer - awaiting approval of new feature key;
putative"
join(203911..203959,204084..204377)
/partial
/genes="TCRBV555"
/notes="putative"
/pseudo
/codon_start=1
<203911..203959
/genes="TCRBV555"
/notes="putative"
/pseudo
/number=1
203960..204083
/genes="TCRBV555"
/notes="5' end no gt consensus splice; putative; does not
fit consensus"
/number=1
204084..204377
/genes="TCRBV555"
/notes="putative"
/pseudo
/number=2
204378..204416
/genes="TCRBV555"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
204378..204384
/notes="RSS_heptamer - awaiting approval of new feature
key"
204385..204407
/notes="RSS_spacer - awaiting approval of new feature key"

```

```

misc_signal 204408..204416
/notes="RSS nonamer - awaiting approval of new feature key"
V_segment join(206647..206695,206828..207121)
/genes="TCRBV1S1"
/notes="putative"
<206647..206695
/genes="TCRBV1S1"
/notes="putative"
/number=1
206696..206827
/genes="TCRBV1S1"
/notes="putative"
/number=1
206828..207121
/genes="TCRBV1S1"
/notes="putative"
/number=2
207122..207160
/genes="TCRBV1S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
207122..207160
/genes="TCRBV1S1"
/notes="RSS heptamer - awaiting approval of new feature
key"
207122..207128
/notes="RSS_heptamer - awaiting approval of new feature
key"
207129..207151
/notes="RSS_spacer - awaiting approval of new feature key"
207129..207151
/notes="RSS_spacer - awaiting approval of new feature key"
207152..207160
/notes="RSS nonamer - awaiting approval of new feature key"
207152..207160
/notes="RSS nonamer - awaiting approval of new feature key"
complement(207825..208060)
/notes="putative"
/rpt_family="Alu"
209626..209890
/notes="putative"
/rpt_family="Alu"
complement(211625..211815)
/notes="putative"
/rpt_family="Alu"
214553..214562
/genes="TCRBV12S4"
/notes="conserved decamer; putative"
join(214638..214686,214793..215087)
/genes="TCRBV12S4"
/notes="putative"
<214638..214686
/genes="TCRBV12S4"
/notes="putative"
/number=1
214687..214792
/genes="TCRBV12S4"
/notes="putative"
/number=1
214793..215087
/genes="TCRBV12S4"

```

```

/misc_signal
/number=2
215088..215126
/clone="TCRBV12S4"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
215088..215094
/notes="RSS_heptamer - awaiting approval of new feature
key"
215095..215117
/notes="RSS_spacer - awaiting approval of new feature key"
215118..215126
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(221378..221646)
/rpt_family="Alu"
/notes="putative"
222287..222296
/clone="TCRBV21S1"
/notes="conserved decamer; putative"
Join(222397..222445,222547..222844)
/clone="TCRBV21S1"
/notes="putative"
<222397..222445
/clone="TCRBV21S1"
/notes="putative"
/number=1
222446..222546
/clone="TCRBV21S1"
/notes="putative"
/number=1
222547..222844
/clone="TCRBV21S1"
/notes="putative"
/number=2
222845..222883
/clone="TCRBV21S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
222845..222851
/notes="RSS_heptamer - awaiting approval of new feature
key"
222852..222874
/notes="RSS_spacer - awaiting approval of new feature key"
222875..222883
/notes="RSS_nonamer - awaiting approval of new feature key"
225419..225732
/notes="putative"
/rpt_family="LINE 1"
complement(228277..228621)
/notes="putative"
/rpt_family="MLTID"
229815..229824
/clone="TCRBV8S4"
/notes="conserved decamer; putative"
Join(229930..229975,230075..230372)
/clone="TCRBV8S4"
/notes="putative"
/pseudo
/codon_start=1
<229930..229975
/clone="TCRBV8S4"
/notes="putative"
/pseudo
/number=1

```

```

intron
229976..230074
/clone="TCRBV8S4"
/notes="putative"
/number=1
230075..230372
/clone="TCRBV8S4"
/notes="putative"
/pseudo
/number=2
230204..230630
/notes="The gap between cosmid clones H18 and V13C15 was
closed by the sequencing of a PCR reaction product."
/organism="Homo sapiens"
/germline
/sequenced_mol="DNA"
/clone="H18/G15gap"
/citation=[1]
230373..230411
/clone="TCRBV8S4"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
230373..230379
/notes="RSS_heptamer - awaiting approval of new feature
key"
230380..230402
/notes="RSS_spacer - awaiting approval of new feature key"
230403..230411
/notes="RSS_nonamer - awaiting approval of new feature key"
230482..230750
/notes="putative"
/rpt_family="Alu"
230528..267156
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="G15"
/clone_lib="Kai Wang's"
/citation=[1]
233955..234049
/notes="putative"
/rpt_family="Alu"
complement(235992..236357)
/notes="putative"
/rpt_family="Alu"
239619..239628
/clone="TCRBV12S3"
/notes="conserved decamer; putative"
Join(239704..239752,239859..240153)
/clone="TCRBV12S3"
/notes="putative"
<239704..239752
/notes="putative"
/clone="TCRBV12S3"
/notes="putative"
/number=1
239753..239858
/clone="TCRBV12S3"
/notes="putative"
/number=1
239859..240153
/clone="TCRBV12S3"
/notes="putative"
/number=2

```

```

misc_signal 240154..240192
/gene="TCRBV12S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
240154..240160
/notes="RSS_heptamer - awaiting approval of new feature
key"
240161..240183
/notes="RSS_spacer - awaiting approval of new feature key"
240184..240192
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit complement(240700..240964)
/notes="putative"
repeat_unit /rpt_family="Alu"
241323..241603
/notes="putative"
repeat_unit /rpt_family="Alu"
243147..243302
/notes="243302; putative"
repeat_unit /rpt_family="Alu"
complement(244346..244691)
/notes="putative"
repeat_unit /rpt_family="THE transposon like element"
complement(246039..246420)
/notes="putative"
repeat_unit /rpt_family="MSTA"
248553..248562
/gene="TCRBV21S3"
/notes="conserved decamer; putative"
V_segment join(248664..248712,248805..249102)
/gene="TCRBV21S3"
/notes="putative"
exon /codon_start=1
<248664..248712
/gene="TCRBV21S3"
/notes="putative"
number=1
join(248664..248712,248805..249102)
/partial
/gene="TCRBV21S3"
/notes="This CDS feature is included to show the
translation of the corresponding V segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624708"
/codon_start=1
/translation="MCTRLLCWALCLLCAELTEAGVAQSPRYKIIEKRSVATWCPN
ISGHATLYWYQOILGGPKLLIQFQNGVDDSLPKDRFSARLKGVDSTLTKQPKK
LEDSAVYLCASSL"
248713..248804
/gene="TCRBV21S3"
/notes="putative"
number=1
exon 248805..249102
/gene="TCRBV21S3"
/notes="putative"
number=2
misc_signal 249103..249141
/gene="TCRBV21S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
249103..249109
/notes="RSS_heptamer - awaiting approval of new feature
key"
249110..249132

```

```

misc_signal 249133..249141
/notes="RSS_spacer - awaiting approval of new feature key"
repeat_unit /note="RSS nonamer - awaiting approval of new feature key"
complement(249302..249573)
/notes="putative"
repeat_unit /rpt_family="Alu"
complement(250307..250649)
/notes="putative"
repeat_unit /rpt_family="THE transposon like element"
252152..252232
/notes="putative"
repeat_unit /rpt_family="MER4"
complement(253936..254362)
/notes="putative"
repeat_unit /rpt_family="MLTID"
255476..255485
/notes="conserved decamer; putative"
V_segment join(255591..255636,255736..256033)
/partial
/gene="TCRBV8S5"
/notes="putative"
pseudo
/codon_start=1
exon <255591..255636
/gene="TCRBV8S5"
/notes="putative"
pseudo
number=1
intron 255637..255735
/gene="TCRBV8S5"
/notes="putative"
number=1
exon 255736..256033
/gene="TCRBV8S5"
/notes="putative"
pseudo
number=2
misc_signal 256034..256072
/gene="TCRBV8S5"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
256034..256040
/notes="RSS_heptamer - awaiting approval of new feature
key"
256041..256063
/notes="RSS_spacer - awaiting approval of new feature key"
256064..256072
/notes="RSS nonamer - awaiting approval of new feature key"
256774..256996
/notes="putative"
repeat_unit /rpt_family="Alu"
257519..2496573
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="X1A"
/clone_lib="Kai Wang's"
/citation={1}
261812..262088
/notes="putative"
repeat_unit /rpt_family="Alu"
263528..263921

```

```

/misc_feature
/rpt family="LINE 1"
265632..265641
/genes="TCRBV13S1"
/notes="conserved decamer; putative"
V_segment
join(265721..265769,265862..266156)
/genes="TCRBV13S1"
/notes="putative"
exon
<265721..265769
/genes="TCRBV13S1"
/notes="putative"
/number=1
intron
265770..265861
/genes="TCRBV13S1"
/notes="putative"
/number=1
exon
265862..266156
/genes="TCRBV13S1"
/notes="putative"
/number=2
/misc_signal
266157..266195
/genes="TCRBV13S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
266157..266163
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
266164..266186
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
266187..266195
/notes="RSS_nonamer - awaiting approval of new feature key"
/misc_feature
269768..269777
/genes="TCRBV6S11"
/notes="conserved decamer; putative"
V_segment
join(269879..269927,270043..270340)
/genes="TCRBV6S11"
/notes="putative"
/number=1
exon
/codon start=1
<269879..269927
/genes="TCRBV6S11"
/notes="putative"
/number=1
CDS
join(269879..269927,270043..270340)
/partial
/genes="TCRBV6S11"
/notes="This CDS feature is included to show the
translation of the corresponding V segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624709"
/codon start=1
translation="MGRILLQWVVLGFLGTDHTGAGVSQSPRYKVAKRGROVALRCDS
ISGHVTLVWYRQTLGGQSEVLTYSQSDAQROKSGRPSGRFSNERPERSVSTIKIQRT
QGUSAVYICASSLA"
intron
269928..270042
/genes="TCRBV6S11"
/notes="putative"
/number=1
exon
270043..270340
/genes="TCRBV6S11"
/notes="putative"
/number=2
/misc_signal
270341..270379
/genes="TCRBV6S11"
/notes="RSS - awaiting approval of new feature key;

```

```

/misc_signal
heptamer is at 5' end; putative"
270341..270347
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
270348..270370
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
270371..270379
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit
276368..276698
/notes="putative"
/number=1
V_segment
/rpt family="MSTA"
join(277826..277874,277997..278290)
/genes="TCRBV5S6"
/notes="putative"
exon
<277826..277874
/genes="TCRBV5S6"
/notes="putative"
/number=1
intron
277875..277996
/genes="TCRBV5S6"
/notes="putative"
/number=1
exon
277997..278290
/genes="TCRBV5S6"
/number=2
/misc_signal
278291..278329
/genes="TCRBV5S6"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
278291..278297
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
278298..278320
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
278321..278329
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit
281194..281439
/notes="putative"
/misc_feature
/rpt family="Alu"
284207..284216
/genes="TCRBV13S6"
/notes="conserved decamer; putative"
V_segment
join(284296..284344,284434..284728)
/genes="TCRBV13S6"
/notes="putative"
exon
<284296..284344
/genes="TCRBV13S6"
/notes="putative"
/number=1
intron
284345..284433
/genes="TCRBV13S6"
/notes="putative"
exon
284434..284728
/genes="TCRBV13S6"
/notes="putative"
/number=2
/misc_signal
284729..284767
/genes="TCRBV13S6"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
284729..284735
/notes="RSS_heptamer - awaiting approval of new feature
key"

```



```

misc_signal 303012..303018
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 303019..303041
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 303042..303050
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit complement(303862..304128)
/notes="putative"
/rpt_family="Alu"
misc_feature 306783..306792
/genes="TCRBV6S4"
/notes="conserved decamer; putative"
V_segment join(306894..306942,307089..307386)
/genes="TCRBV6S4"
/notes="putative"
exon <306894..306942
/genes="TCRBV6S4"
/notes="putative"
/number=1
intron 306943..307088
/genes="TCRBV6S4"
/notes="putative"
/number=1
exon 307089..307386
/genes="TCRBV6S4"
/notes="putative"
/number=2
misc_signal 307387..307425
/genes="TCRBV6S4"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 307387..307393
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 307394..307416
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 307417..307425
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit 313325..313655
/notes="putative"
/rpt_family="MSTA"
misc_feature 314667..314676
/genes="TCRBV5S2"
/notes="conserved decamer; putative"
V_segment join(314783..314831,314955..315248)
/genes="TCRBV5S2"
/notes="putative"
<314783..314831
/genes="TCRBV5S2"
/notes="putative"
/number=1
intron 314832..314954
/genes="TCRBV5S2"
/notes="putative"
/number=1
exon 314955..315248
/genes="TCRBV5S2"
/notes="putative"
/number=2
misc_signal 315249..315287
/genes="TCRBV5S2"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"

```

```

misc_signal 315249..315255
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 315256..315278
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 315279..315287
/notes="RSS_nonamer - awaiting approval of new feature key"
source 321165..333651
/notes=" (vector PME15A) "
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="A2I2partial"
/clone_lib="YAC 234 A6F6"
322006..322015
/genes="TCRBV13S7"
/notes="conserved decamer; putative"
join(322095..322143,322232..322523)
/misc_feature
/number=1
V_segment
/genes="TCRBV13S7"
/notes="putative"
<322095..322143
/genes="TCRBV13S7"
/notes="putative"
/number=1
intron 322144..322231
/genes="TCRBV13S7"
/notes="putative"
/number=1
exon 322232..322523
/genes="TCRBV13S7"
/notes="putative"
/number=2
misc_signal 322524..322562
/genes="TCRBV13S7"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 322524..322530
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 322531..322553
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 322554..322562
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit complement(323316..323575)
/notes="putative"
/rpt_family="Alu"
misc_feature 326226..326235
/genes="TCRBV6S14"
/notes="conserved decamer; putative"
V_segment join(326337..326385,326541..326838)
/genes="TCRBV6S14"
/notes="putative"
<326337..326385
/genes="TCRBV6S14"
/notes="putative"
/number=1
intron 326386..326540
/genes="TCRBV6S14"
/notes="putative"
/number=1
exon 326541..326838
/genes="TCRBV6S14"

```



```

/ gene="TCRBV23S1"
/ note="putative"
/ number=1
370845..371139
/ gene="TCRBV23S1"
/ note="putative"
/ number=2
371140..371178
/ gene="TCRBV23S1"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
371140..371146
/ note="RSS_heptamer - awaiting approval of new feature
key"
371147..371169
/ note="RSS_spacer - awaiting approval of new feature key"
371170..371178
/ note="RSS nonamer - awaiting approval of new feature key"
complement(371174..371631)
/ note="putative"
/ rpt_family="MLT1B"
complement(371186..372141)
/ note="putative"
/ rpt_family="Alu"
complement(373018..373274)
/ note="putative"
/ rpt_family="Alu"
375722..375975
/ note="putative"
/ rpt_family="Alu"
complement(377891..378378)
/ note="putative"
/ rpt_family="MLT2CA"
378984..378993
/ gene="TCRBV12S2"
/ note="conserved decamer; putative"
join(379068..379116,379223..379517)
/ gene="TCRBV12S2"
/ note="putative"
<379068..379116
/ gene="TCRBV12S2"
/ note="putative"
/ number=1
379117..379222
/ gene="TCRBV12S2"
/ note="putative"
/ number=1
379223..379517
/ gene="TCRBV12S2"
/ note="putative"
/ number=2
379518..379556
/ gene="TCRBV12S2"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
379518..379524
/ note="RSS_heptamer - awaiting approval of new feature
key"
379525..379547
/ note="RSS_spacer - awaiting approval of new feature key"
379548..379556
/ note="RSS nonamer - awaiting approval of new feature key"
380360..380626
/ note="putative"

```

```

/ rpt_family="Alu"
380936..381190
/ note="putative"
/ rpt_family="Alu"
381368..420906
/ organisms="Homo sapiens"
/ germline
/ sequenced_mol="DNA"
/ clone="H12.18"
/ citations=[2]
382199..382460
/ note="putative"
/ rpt_family="Alu"
382674..383012
/ note="putative"
/ rpt_family="MSTA"
complement(383847..384274)
/ note="putative"
/ rpt_family="LINE 1"
384754..385422
/ note="putative"
/ rpt_family="LINE 1"
389605..389615
/ gene="TCRBV21S4"
/ note="conserved decamer; putative"
join(389716..389764,389856..390153)
/ gene="TCRBV21S4"
/ note="putative"
<389716..389764
/ gene="TCRBV21S4"
/ note="putative"
/ number=1
389765..389855
/ gene="TCRBV21S4"
/ note="putative"
/ number=1
389856..390153
/ gene="TCRBV21S4"
/ note="putative"
/ number=2
390154..390192
/ gene="TCRBV21S4"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
390154..390160
/ note="RSS_heptamer - awaiting approval of new feature
key"
390161..390183
/ note="RSS_spacer - awaiting approval of new feature key"
390184..390192
/ note="RSS nonamer - awaiting approval of new feature key"
complement(393711..394157)
/ note="putative"
/ rpt_family="MLT1D"
395203..395212
/ gene="TCRBV8S1"
/ note="conserved decamer; putative"
join(395318..395366,395467..395764)
/ gene="TCRBV8S1"
/ note="putative"
<395318..395366
/ gene="TCRBV8S1"
/ note="putative"
/ number=1

```

```

intron      395367..395466
            /gene="TCRBV8S1"
            /note="putative"
            /number=1
exon        395467..395764
            /gene="TCRBV8S1"
            /note="putative"
            /number=2
misc_signal 395765..395803
            /gene="TCRBV8S1"
            /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            395765..395771
            /note="RSS heptamer - awaiting approval of new feature
            key"
            395772..395794
            /note="RSS spacer - awaiting approval of new feature key"
            395795..395803
            /note="RSS nonamer - awaiting approval of new feature key"
            complement(397023..397460)
            /note="putative"
            /rpt_family="MLTID"
            398526..398535
            /gene="TCRBV8S2"
            /note="conserved decamer; putative"
            join(398641..398689,398790..399087)
            /gene="TCRBV8S2"
            /note="putative"
            <398641..398689
            /gene="TCRBV8S2"
            /note="putative"
            /number=1
            398690..398789
            /gene="TCRBV8S2"
            /note="putative"
            /number=1
            398790..399087
            /gene="TCRBV8S2"
            /note="putative"
            /number=2
            399088..399126
            /gene="TCRBV8S2"
            /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            399088..399094
            /note="RSS heptamer - awaiting approval of new feature
            key"
            399095..399117
            /note="RSS spacer - awaiting approval of new feature key"
            399118..399126
            /note="RSS nonamer - awaiting approval of new feature key"
            402029..402832
            /note="putative"
            /rpt_family="HUMERSP2A"
            complement(403915..404184)
            /note="putative"
            404140..442094
            /rpt_family="Alu"
            /organism="Homo sapiens"
            /germline
            /sequenced_mol="DNA"
            /clone="H130"
            /citation={2}
            405472..409693
            repeat_unit

```

```

            /note="putative"
            /rpt_family="LINE 1"
            409045..409306
            /note="putative"
            /rpt_family="Alu"
            complement(409793..410042)
            /note="putative"
            /rpt_family="Alu"
            410544..411352
            /note="putative"
            /rpt_family="HUMERVKB1"
            411803..412066
            /note="putative"
            /rpt_family="Alu"
            412870..413127
            /note="putative"
            /rpt_family="Alu"
            415714..415723
            /gene="TCRBV8S3"
            /note="conserved decamer; putative"
            join(415832..415880,415981..416278)
            /gene="TCRBV8S3"
            /note="putative"
            <415832..415880
            /gene="TCRBV8S3"
            /note="putative"
            /number=1
            415881..415980
            /gene="TCRBV8S3"
            /note="putative"
            /number=1
            415981..416278
            /gene="TCRBV8S3"
            /note="putative"
            /number=2
            416279..416317
            /gene="TCRBV8S3"
            /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            416279..416285
            /note="RSS heptamer - awaiting approval of new feature
            key"
            416286..416308
            /note="RSS spacer - awaiting approval of new feature key"
            416309..416317
            /note="RSS nonamer - awaiting approval of new feature key"
            420729..421157
            /note="putative"
            /rpt_family="MLTID"
            422664..422673
            /gene="TCRBV16S1"
            /note="conserved decamer; putative"
            join(422775..422823,422910..423207)
            /gene="TCRBV16S1"
            /note="putative"
            <422775..422823
            /gene="TCRBV16S1"
            /note="putative"
            /number=1
            422824..422909
            /gene="TCRBV16S1"
            /note="putative"
            /number=1
            422910..423207
            exon

```

```

/misc_signal /gene="TCRBV16S1"
/number=2
423208..423246
/notes="TCRBV16S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal 423208..423214
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal 423215..423237
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal 423238..423246
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit complement(423257..424217)
/notes="putative"
/misc_feature /rpt_family="Alu"
427725..427734
/genes="TCRBV24S1"
/notes="conserved decamer; putative"
V_segment join(427850..427898,428024..428318)
/genes="TCRBV24S1"
/notes="putative"
exon <427850..427898
/genes="TCRBV24S1"
/notes="putative"
intron /number=1
427899..428023
/genes="TCRBV24S1"
/notes="putative"
exon 428024..428318
/genes="TCRBV24S1"
/notes="putative"
/misc_signal 428319..428357
/genes="TCRBV24S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal 428319..428325
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal 428326..428348
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal 428349..428357
/notes="RSS_nonamer - awaiting approval of new feature key"
/misc_feature 432763..432772
/notes="conserved decamer; putative"
V_segment join(432859..432907,433015..433312)
/partial
/genes="TCRBV25S1"
/notes="putative"
exon /pseudo
<432859..432907
/genes="TCRBV25S1"
/notes="putative"
intron /pseudo
432908..433014
/genes="TCRBV25S1"
/notes="putative"
exon /number=1
433015..433312
/genes="TCRBV25S1"

```

```

/misc_signal /note="putative"
/pseudo
/number=2
433313..433351
/genes="TCRBV25S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal 433313..433319
/genes="TCRBV25S1"
/notes="RSS_heptamer - awaiting approval of new feature
key; putative"
/misc_signal 433320..433342
/genes="TCRBV25S1"
/notes="RSS_spacer - awaiting approval of new feature key;
putative"
/misc_signal 433343..433351
/notes="RSS_nonamer - awaiting approval of new feature key"
/misc_signal 433343..433349
/notes="RSS_nonamer - awaiting approval of new feature key;
putative"
repeat_unit 433523..433794
/notes="putative"
/notes="putative"
repeat_unit /rpt_family="Alu"
434353..434613
/notes="putative"
repeat_unit /rpt_family="Alu"
435607..435864
/notes="putative"
/misc_feature /rpt_family="Alu"
436423..436433
/genes="TCRBV26S1"
/notes="conserved decamer; putative"
V_segment join(436475..436523,436915..437209)
/genes="TCRBV26S1"
/notes="putative"
exon /pseudo
/codon_start=1
<436475..436523
/genes="TCRBV26S1"
/notes="putative"
intron /pseudo
/number=1
436524..436914
/genes="TCRBV26S1"
/notes="putative"
repeat_unit /number=1
436572..436830
/notes="putative"
/notes="putative"
exon /rpt_family="Alu"
436915..437209
/genes="TCRBV26S1"
/notes="putative"
/misc_signal /pseudo
/number=2
437210..437248
/genes="TCRBV26S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal 437210..437216
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal 437217..437239
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal 437240..437248

```

```
source /note="RSS nonamer - awaiting approval of new feature key"
438704..472940
/note="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CCM1"
/germline
/sequenced_mol="DNA"
/clone="Alf"
/clone_lib="YAC 234 A6F6"
/citation=[1]
replace(439095,"a")
/note="from cosmid A16; putative"
replace(439470..439471,"cctt")
/note="from cosmid A16; putative"
replace(439591,"t")
/note="from cosmid A16; putative"
440365..440622
/note="putative"
/rpt_family="Alu"
replace(440424,"c")
/note="from cosmid A16; putative"
replace(440641..440642,"caataata")
/note="from cosmid A16; putative"
replace(440936..440939,"agtg")
/note="from cosmid A16; putative"
replace(441417,"c")
/note="from cosmid A16; putative"
replace(441798..441799,"taa")
/note="from cosmid A16; putative"
441863..442125
/note="putative"
/rpt_family="Alu"
442182..442207
/note="polymorphic"
/rpt_family="microsatellite"
/rpt_unit=442182..442183
/evidence=experimental
complement(442683..442948)
/note="putative"
/rpt_family="Alu"
444936..444983
/note="polymorphic; putative"
/rpt_family="microsatellite"
/rpt_unit=444936..444937
448412..448665
/note="putative"
/rpt_family="Alu"
449251..449509
/note="putative"
/rpt_family="Alu"
449974..450226
/note="putative"
/rpt_family="Alu"
450430..450439
/genes="TCRBV18S1"
/note="conserved decamer; putative"
Join(450546..450594,450867..451164)
/genes="TCRBV18S1"
/note="heptamer is at 5' end; putative"
<450546..450594
/genes="TCRBV18S1"
/note="putative"
/number=1
450595..450866
intron
```

```
/genes="TCRBV18S1"
/note="putative"
/number=1
450867..451164
/genes="TCRBV18S1"
/note="putative"
/number=2
451165..451203
/genes="TCRBV18S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
451165..451171
/note="RSS heptamer - awaiting approval of new feature
key"
451172..451194
/note="RSS spacer - awaiting approval of new feature key"
451195..451203
/note="RSS nonamer - awaiting approval of new feature key"
452324..486943
/note="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CCM1"
/germline
/sequenced_mol="DNA"
/clone="C215"
/clone_lib="YAC 234 A72B3"
/citation=[1]
Join(453806..453854,453987..454281)
/genes="TCRBV17S1"
/note="putative"
<453806..453854
/genes="TCRBV17S1"
/note="putative"
/number=1
453987..454281
/genes="TCRBV17S1"
/note="putative"
/number=2
454282..454320
/genes="TCRBV17S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
454282..454288
/note="RSS heptamer - awaiting approval of new feature
key"
454289..454311
/note="RSS spacer - awaiting approval of new feature key"
454312..454320
/note="RSS nonamer - awaiting approval of new feature key"
complement(454622..454883)
/note="putative"
/rpt_family="Alu"
457907..>457907
/note="translocation breakpoint; t(7:9); putative"
complement(461020..461281)
/note="putative"
/rpt_family="Alu"
Join(461476..461506,461845..462148)
/genes="TCRBV2S1"
/note="putative"
V_segment
```

```

exon
<461476..461506
/feature="TCRBV251"
/feature="putative"
/number=1
461507..461844
/feature="TCRBV251"
/feature="putative"
/number=1
461845..462148
/feature="TCRBV251"
/number=2
462149..462187
/feature="TCRBV251"
/feature="RSS" - awaiting approval of new feature key;
heptamer is at 5' end; putative"
462149..462155
/feature="TCRBV251"
/feature="RSS_heptamer" - awaiting approval of new feature
key; putative"
462156..462178
/feature="TCRBV251"
/feature="RSS_spacer" - awaiting approval of new feature key;
putative"
462179..462187
/feature="TCRBV251"
/feature="RSS_nonamer" - awaiting approval of new feature key;
putative"
465204..465245
/feature="polymorphic; putative"
/rpt_family="microsatellite"
/rpt_unit=465204..465205
468059..468314
/feature="putative"
/rpt_family="Alu"
468548..468804
/feature="putative"
/rpt_family="Alu"
470316..470574
/feature="putative"
/rpt_family="Alu"
470593..470624
/feature="polymorphic; putative"
/rpt_family="microsatellite"
/rpt_unit=470593..470596
470746..471356
/feature="putative"
/rpt_family="LINE 1"
471558..471567
/feature="TCRBV10S1"
/feature="conserved decamer; putative"
join(471559..471708,471822..472119)
/feature="TCRBV10S1"
/feature="putative"
/number=1
471659..471708
/partial
/feature="TCRBV10S1"
/feature="putative"
/number=1
471709..471821
/feature="TCRBV10S1"
/feature="putative"
/number=1

```

```

exon
471822..472119
/feature="TCRBV10S1"
/feature="putative"
/number=2
472120..472158
/feature="TCRBV10S1"
/feature="RSS" - awaiting approval of new feature key;
heptamer is at 5' end; putative"
472120..472126
/feature="RSS_heptamer" - awaiting approval of new feature
key"
472127..472149
/feature="RSS_spacer" - awaiting approval of new feature key"
472150..472158
/feature="RSS_nonamer" - awaiting approval of new feature key"
complement(472732..472994)
/feature="putative"
/rpt_family="Alu"
complement(475772..475851)
/feature="putative"
/rpt_family="MIR"
476399..476408
/feature="TCRBV29S1"
/feature="conserved decamer; putative"
join(476480..476528,476661..476730)
/feature="TCRBV29S1"
/feature="putative"
/number=1
476529..476660
/feature="TCRBV29S1"
/feature="putative"
/number=1
476661..476730
/feature="TCRBV29S1"
/feature="putative"
/number=2
476931..476969
/feature="TCRBV29S1"
/feature="RSS" - awaiting approval of new feature key;
heptamer is at 5' end; putative"
476931..476937
/feature="RSS_heptamer" - awaiting approval of new feature
key"
476938..476960
/feature="RSS_spacer" - awaiting approval of new feature key"
476961..476969
/feature="RSS_nonamer" - awaiting approval of new feature key"
480626..480635
/feature="TCRBV19S1"
/feature="conserved decamer; putative"
join(480699..480747,480898..481195)
/feature="TCRBV19S1"
/feature="putative"
/number=1
480699..480747
/feature="TCRBV19S1"
/feature="putative"

```

```

/misc_signal
/number=1
480748..480897
/genes="TCRBV19S1"
/notes="lacks functional splice donor; putative; does not
fit consensus"
/number=1
480898..481195
/genes="TCRBV19S1"
/notes="putative"
/pseudo
/number=2
481196..481234
/genes="TCRBV19S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
481196..481202
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
481203..481225
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
481226..481234
/notes="RSS_nonamer - awaiting approval of new feature key"
source
482597..515832
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="ATCC 1521"
/germline
/sequenced_mol="DNA"
/clone="G1"
/clone_lib="Kai Wang's"
/citation=[1]
replace(484618,"c")
/notes="from cosmid G1; putative"
485079..487500
/notes="putative"
/rpt_family="LINE 1"
488729..489015
/notes="putative"
/rpt_family="HUMERSP2A"
489941..489950
/genes="TCRBV15S1"
/notes="conserved decamer; putative"
Join(490040..490088,490221..490516)
/genes="TCRBV15S1"
/notes="putative"
<490040..490088
/genes="TCRBV15S1"
/notes="putative"
/number=1
490089..490220
/genes="TCRBV15S1"
/notes="putative"
/number=1
490221..490516
/genes="TCRBV15S1"
/notes="putative"
/number=2
490517..490555
/genes="TCRBV15S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
490517..490523
/notes="RSS_heptamer - awaiting approval of new feature

```

```

/misc_signal
490524..490546
/notes="RSS_spacer - awaiting approval of new feature key"
490547..490555
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(493122..493784)
/notes="putative"
/rpt_family="LINE 1"
complement(493240..493514)
/notes="putative"
/rpt_family="Alu"
complement(494073..495301)
/notes="putative"
/rpt_family="LINE 1"
complement(497524..498018)
/notes="putative"
/rpt_family="MER2"
501895..502102
/notes="putative"
504326..504335
/genes="TCRBV11S1"
/notes="conserved decamer; putative"
Join(504415..504463,504588..504882)
/genes="TCRBV11S1"
/notes="putative"
504415..504463
/partial
/genes="TCRBV11S1"
/notes="putative"
/number=1
504464..504587
/genes="TCRBV11S1"
/notes="putative"
/number=1
504588..504882
/genes="TCRBV11S1"
/notes="putative"
/number=2
504707..538564
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="C68"
/clone_lib="YAC 199 B157G9"
/citation=[1]
504883..504921
/genes="TCRBV11S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
504883..504889
/notes="RSS_heptamer - awaiting approval of new feature
key"
504890..504912
/notes="RSS_spacer - awaiting approval of new feature key"
504913..504921
/notes="RSS_nonamer - awaiting approval of new feature key"
506376..506604
/notes="putative"
/rpt_family="MER2"
507025..507859
/notes="putative"

```

```

repeat_unit      /rpt_family="LINE 1"
                  complement(508000..510053)
                  /note="putative"
variation         /rpt_family="LINE 1"
                  replace(508199,"g")
                  /note="from cosmid C68; putative"
repeat_unit      508498..508759
                  /note="putative"
variation         /rpt_family="Alu"
                  replace(510841..510843,"ct")
                  /note="from cosmid C68; putative"
repeat_unit      complement(510867..511137)
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  512915..514739
                  /note="putative"
repeat_unit      /rpt_family="LINE 1"
                  513363..513884
                  /note="putative"
variation         /rpt_family="MLT2A"
                  replace(513888,"c")
                  /note="from cosmid C68; putative"
V_segment        join(515044..515090,515129..515497)
                  /gene="TCRBV33S1"
                  /note="putative"
                  /pseudo
exon              <515044..515090
                  /gene="TCRBV33S1"
                  /note="putative"
                  /pseudo
                  /number=1
intron            /codon_start=1
                  515091..515128
                  /gene="TCRBV33S1"
                  /note="putative"
                  /number=1
exon              515129..515497
                  /gene="TCRBV33S1"
                  /note="putative"
                  /pseudo
                  /number=2
misc_signal       /codon_start=1
                  515498..515536
                  /gene="TCRBV33S1"
                  /note="RSS - awaiting approval of new feature key;
                  heptamer is at 5' end; putative"
misc_signal       515498..515504
                  /note="RSS_heptamer - awaiting approval of new feature
                  key"
misc_signal       515505..515527
                  /note="RSS_spacer - awaiting approval of new feature key"
misc_signal       515528..515536
                  /note="RSS nonamer - awaiting approval of new feature key"
repeat_unit      518289..518549
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  complement(520882..521434)
                  /note="putative"
repeat_unit      /rpt_family="LINE 1"
                  523833..524090
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  524764..525025
                  /note="putative"

```

```

V_segment        /rpt_family="Alu"
                  join(529336..529384,529526..529820)
                  /gene="TCRBV28S1"
                  /note="putative"
                  /pseudo
exon              <529336..529384
                  /gene="TCRBV28S1"
                  /note="putative"
                  /pseudo
                  /number=1
intron            /codon_start=1
                  529385..529525
                  /gene="TCRBV28S1"
                  /note="putative"
                  /number=1
exon              529526..529820
                  /gene="TCRBV28S1"
                  /note="putative"
                  /pseudo
                  /number=2
misc_signal       /codon_start=1
                  529821..529859
                  /gene="TCRBV28S1"
                  /note="RSS - awaiting approval of new feature key;
                  heptamer is at 5' end; putative"
                  529821..529827
                  /gene="TCRBV28S1"
                  /note="RSS_heptamer - awaiting approval of new feature
                  key; putative"
                  529828..529850
                  /gene="TCRBV28S1"
                  /note="RSS_spacer - awaiting approval of new feature key;
                  putative"
                  529851..529859
                  /gene="TCRBV28S1"
                  /note="RSS_nonamer - awaiting approval of new feature key;
                  putative"
                  530024..530166
                  /note="putative"
                  /rpt_family="MER4"
                  complement(530975..531036)
                  /note="putative"
                  /rpt_family="MLT1A"
                  531638..565300
                  /note="(vector PWE15A)"
                  /organism="Homo sapiens"
                  /cell_line="CGM1"
                  /germline
                  /sequenced_mol="DNA"
                  /clone="C21"
                  /clone_lib="YAC 234 A72B3"
                  /citation=[1]
                  replace(531984..531985,"gcc")
                  /note="found in cosmid C21; putative"
                  replace(533511..533512,"tgg")
                  /note="found in cosmid C21; putative"
                  533983..534356
                  /note="putative"
                  /rpt_family="THE transposon like element"
                  535432..535690
                  /note="putative"
                  /rpt_family="Alu"
                  536867..537285

```

```

/note="putative"
/rpt_family="MLTID"
replace(537605..537607,"ta")
/note="found in cosmid C21; putative"
539683..539720
/note="polymorphic; putative"
/rpt_family="microsatellite"
/rpt_unit=539683..539684
544636..544866
/notes="putative"
/rpt_family="MLTIB"
join(545039..545080,545214..545572)
/genes="TCRBV34S1"
/notes="putative"
/pseudo
/codon_start=1
<545039..545080
/genes="TCRBV34S1"
/notes="putative"
/pseudo
/number=1
/codon_start=1
545081..545213
/genes="TCRBV34S1"
/notes="putative"
/number=1
545214..545572
/genes="TCRBV34S1"
/notes="putative"
/pseudo
/number=2
/codon_start=1
545573..545579
/genes="TCRBV34S1"
/notes="RSS - awaiting approval of new feature key;
degraded RSS; putative"
545573..545579
/genes="TCRBV34S1"
/notes="RSS heptamer - awaiting approval of new feature
key; putative"
complement(545711..546155)
/notes="putative"
/rpt_family="MLTID"
548386..548634
/notes="putative"
/rpt_family="Alu"
548863..548872
/genes="TCRBV14S1"
/notes="conserved decamer; putative"
join(549038..549086,549216..549510)
/genes="TCRBV14S1"
/notes="putative"
<549038..549086
/genes="TCRBV14S1"
/notes="putative"
/number=1
549087..549215
/genes="TCRBV14S1"
/notes="putative"
/number=1
549216..549510
/genes="TCRBV14S1"
/notes="putative"
/number=2

```

```

misc_signal 549511..549549
/genes="TCRBV14S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
549511..549549
/genes="TCRBV14S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
549511..549517
/notes="RSS heptamer - awaiting approval of new feature
key"
549518..549540
/notes="RSS spacer - awaiting approval of new feature key"
549541..549549
/notes="RSS nonamer - awaiting approval of new feature key"
551440..551700
/notes="putative"
/rpt_family="Alu"
552491..590912
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="ATCC 1521"
/germline
/sequenced_mol="DNA"
/clone="X11"
/clone_lib="Kai Wang's"
/citation=[1]
553554..553886
/notes="putative"
/rpt_family="MLTIB"
554259..554268
/genes="TCRBV3S1"
/notes="conserved decamer; putative"
join(554328..554376,554514..554808)
/partial
/genes="TCRBV3S1"
/notes="This CDS feature is included to show the
translation of the corresponding V segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624710"
/codon_start=1
/translation="MGIIRLLCRVAFCTIAGLVGVKVTQSSRYLVKRTGKVFLECVQ
DMDHENMFWRQDPGLGIRLLYFSYDVKKKEKGDIPGYSVSREKRFSLLESAST
NQTSMYLCASSL"
join(554328..554376,554514..554808)
/genes="TCRBV3S1"
/notes="putative"
<554328..554376
/genes="TCRBV3S1"
/notes="putative"
/number=1
554377..554513
/genes="TCRBV3S1"
/notes="putative"
/number=1
replace(554495,"a")
/notes="found in cosmid X11; putative"
554514..554808
/genes="TCRBV3S1"
/notes="putative"
/number=2
554809..554847
/genes="TCRBV3S1"

```



```

misc_signal      /note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
554809..554815
/note="RSS_heptamer - awaiting approval of new feature
key"
554816..554838
/note="RSS_spacer - awaiting approval of new feature key"
554839..554847
/note="RSS nonamer - awaiting approval of new feature key"
complement(557145..557400)
/note="putative"
/rpt family="Alu"
557655..558216
/note="putative"
/rpt family="LINE1"
559309..561579
/note="putative"
/rpt family="THE1 transposon like element"
replace(561498,"a")
/note="found in cosmid X11; putative"
replace(561742,"g")
/note="found in cosmid X11; putative"
replace(562192,"c")
/note="from cosmid X11; putative"
replace(562232,"a")
/note="from cosmid X11; putative"
replace(562694,"a")
/note="from cosmid X11; putative"
replace(563416,"c")
/note="from cosmid X11; putative"
replace(563871,"t")
/note="from cosmid X11; putative"
replace(564119,"t")
/note="from cosmid X11; putative"
565220
/note="The 3' breakpoint occurs at basepair 592356,
602910, 613477, or 623282 -- identical positions within a
homology unit that is tandemly repeated.; translocation
breakpoint t(7:9); putative"
567741..568016
/note="putative"
/rpt family="Alu"
567946..569865
/note="putative"
/rpt family="LINE 1"
568023..569813
/note="putative"
/rpt family="LINE1"
569393..606124
/note="(vector PME15A)"
/organism="Homo sapiens"
/germline
/sequenced_mol="DNA"
/clone="X6a"
/clone_lib="Kai Wang's"
/tissue_lib="ATCC 1521"
/citation=[1]
replace(570855,"c")
/note="from cosmid X6A; putative"
replace(570966..570968,"aa")
/note="from cosmid X6A; putative"
replace(572279,"a")
/note="from cosmid X6A; putative"
replace(573140,"a")

```

```

variation      /note="from cosmid X6A; putative"
replace(573457,"a")
/note="from cosmid X6A; putative"
replace(573462,"a")
/note="from cosmid X6A; putative"
join(573946..573979,574259..574559)
/genes="TCRBV4S1"
/note="putative"
/codon_start=1
<573946..573979
/genes="TCRBV4S1"
/note="putative"
/numbers=1
join(573946..573979,574259..574559)
/partial
/genes="TCRBV4S1"
/note="This CDS feature is included to show the
translation of the corresponding V_segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624711"
/codon_start=1
/translation="MLSLLILLGLGVFSAVISQKPSRDICQGTSLTIQCVDQSQV
TMFWVRQPCQSLTIATANQCSSEATYSGFVDKPFISRPNLTFSLITVSNMSPED
SSIIYLCGVE"
573980..574258
/genes="TCRBV4S1"
/note="putative"
/numbers=1
replace(574024..574025,"gggtggaag")
/note="from cosmid X6A; putative"
574259..574559
/genes="TCRBV4S1"
/note="putative"
/number=2
574560..574598
/genes="TCRBV4S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
574560..574566
/note="RSS_heptamer - awaiting approval of new feature
key"
574567..574589
/note="RSS_spacer - awaiting approval of new feature key"
574590..574598
/note="RSS nonamer - awaiting approval of new feature key"
replace(575119,"c")
/note="from cosmid X6A; putative"
replace(578315..578317,"ca")
/note="from cosmid X6A; putative"
replace(578641,"c")
/note="from cosmid X6A; putative"
replace(579597,"t")
/note="from cosmid X6A; putative"
replace(579607,"c")
/note="from cosmid X6A; putative"
replace(579794,"c")
/note="from cosmid X6A; putative"
replace(579842..579843,"aaga")
/note="from cosmid X6A; putative"
580170..580273
/note="putative"
/rpt family="MIR1"
complement(580171..580256)
/note="putative"

```

```

variation      /rpt_family="MIR"
repeat_unit    replace(580374,"g")
               /note="from cosmid X6A; putative"
               580376..580477
               /note="putative"
variation      /rpt_family="MIR"
               replace(580467,"g")
               /note="from cosmid X6A; putative"
variation      replace(580471,"ca")
               /note="from cosmid X6A; putative"
variation      replace(581770,"a")
               /note="from cosmid X6A; putative"
variation      replace(582745,"t")
               /note="from cosmid X6A; putative"
variation      replace(582949,"a")
               /note="from cosmid X6A; putative"
variation      replace(583153..583192,584223..584382,585442..585695,
mRNA           586099..586235,586536..586688)
               /gene="TRYA"
exon           <583153..583192
               /gene="TRYA"
               /note="putative"
               /number=1
CDS            join(583153..583192,584223..584382,585442..585695,
               586099..586235,586536..586688)
               /gene="TRYA"
               /note="putative; NCBI gi: 624712"
               /codon_start=1
               /product="trypsinogen A"
               /translation="MNPILLITFVAALAAPFDDDDKIVGCVNCEENSVYQVSLNSG
               YHF CGGLINEQVMVSGHCYKSRISQVRIGEHNIIEVLENEQF INAAKIIIRHPQYDRK
               TLNIDIMLKLSRAVIRVSTISLTAPPATGTGKLISGNGNTASSGADYPDELQC
               LDAPVLSQAKCEASVPKITSNNFCVGTLEGCKSCQGDGPGPVVCGQLQGVNSWGD
               GCAQRKPGVYTKVNTVYRWIKNTTAANS"
               583193..584222
               /gene="TRYA"
               /note="putative; does not fit consensus"
               /number=1
variation      replace(583491,"a")
exon           /note="from cosmid X6A; putative"
               584223..584382
               /gene="TRYA"
               /number=2
intron         584383..585441
               /gene="TRYA"
               /note="putative"
               /number=2
variation      replace(585107,"c")
exon           /note="from cosmid X6A; putative"
               585442..585695
               /gene="TRYA"
               /note="putative"
               /number=3
intron         585696..586098
               /gene="TRYA"
               /note="putative"
               /number=3
exon           586099..586235
               /gene="TRYA"
               /note="putative"
               /number=4
variation      replace(586130,"t")
               /gene="TRYA"
               /note="from cosmid X6A; putative"

```

```

intron         586236..586535
               /gene="TRYA"
               /note="putative"
               /number=4
exon           586536..586688
               /gene="TRYA"
               /note="putative"
               /number=5
variation      replace(586682,"t")
               /gene="TRYA"
               /note="from cosmid X6A; putative"
               586719..586724
               /note="putative"
               replace(587095,"t")
               /note="from cosmid X6A; putative"
               replace(587353,"g")
               /note="from cosmid X6A; putative"
               replace(587365,"a")
               /note="from cosmid X6A; putative"
               replace(588389,"a")
               /note="from cosmid X6A; putative"
               replace(588490,"c")
               /note="from cosmid X6A; putative"
               replace(588898,"t")
               /note="from cosmid X6A; putative"
               complement(589057..589313)
               /note="putative"
               /rpt_family="Alu"
               replace(589124,"c")
               /note="from cosmid X6A; putative"
               589531..590009
               /note="putative"
               /rpt_family="LINE1"
               replace(589535,"c")
               /note="from cosmid X6A; putative"
               replace(589738,"c")
               /note="from cosmid X6A; putative"
               replace(589740,"t")
               /note="from cosmid X6A; putative"
               replace(590011..590015,at")
               /note="from cosmid X6A; putative"
               replace(590085,"g")
               /note="from cosmid X6A; putative"
               590218..590472
               /note="putative"
               /rpt_family="Alu"
               590287..590435
               /note="putative"
               /rpt_family="Alu"
               replace(590308,"g")
               /note="from cosmid X6A; putative"
               replace(590334,"c")
               /note="from cosmid X6A; putative"
               replace(590506,"a")
               /note="from cosmid X6A; putative"
               replace(590509..590510,atg")
               /note="from cosmid X6A; putative"
               replace(591483,"t")
               /note="from cosmid X6A; putative"
               replace(592143,"a")
               /note="from cosmid X6A; putative"
               replace(592154,"a")
               /note="from cosmid X6A; putative"
               join(<594083..594122,595142..595301,596361..596614,
mRNA

```

exon
 597022..597158,597460..597610)
 /gene="TRYB"
 /note="putative"
 <594083..594122
 /gene="TRYB"
 /note="putative"
 /pseudo
 /number=1
 join(594083..594122,595142..595301,596361..596614,
 597022..597158,597460..597610)
 /gene="TRYB"
 /note="putative"
 /pseudo
 /product="trypsinogen B"
 594123..595141
 /partial
 /gene="TRYB"
 /note="putative"
 /number=1
 595142..595301
 /gene="TRYB"
 /note="putative"
 /pseudo
 /number=2
 595302..596360
 /gene="TRYB"
 /note="putative"
 /number=2
 596361..596614
 /gene="TRYB"
 /note="putative"
 /pseudo
 /number=3
 596615..597021
 /gene="TRYB"
 /note="putative"
 /number=3
 597022..597158
 /gene="TRYB"
 /note="putative"
 /pseudo
 /number=4
 597159..597459
 /partial
 /gene="TRYB"
 /note="putative"
 /number=4
 597460..597610
 /partial
 /gene="TRYB"
 /note="putative"
 /pseudo
 /number=5
 597643..597648
 /gene="TRYB"
 /note="putative"
 complement(599961..600205)
 /note="putative"
 /rpt_family="Alu"
 complement(600520..600879)
 /note="putative"
 /rpt_family="LINE 1"
 join(604575..604698,605728..605887,606946..607199,
 607595..607731,608031..608181)

/partial
 /gene="TRYC"
 /note="putative"
 604575..604698
 /partial
 /gene="TRYC"
 /note="putative"
 /number=1
 join(604659..604698,605728..605887,606946..607199,
 607595..607731,608031..608183)
 /gene="TRYC"
 /note="putative; NCBI gi: 624713"
 /codon_start=1
 /product="trypsinogen C"
 /translation="MNPILLILAFVGAAVFDDDDKIVGGYTCENSVYQVSIASG
 SHEFGGSLISEQWVVSAGHCYKPHIQVRLGEHNIIEVLENEQFINAAKTIIRPKYRI
 TLNDIMLIKLSPTPAVINAHVSTIISLTPAPAGTECLISGNTLSSGADYDELQC
 LDAPVLTQAKCKASYPKITSKMF CVGFLEGGKDCQGGSGGPVVCNGQLAQIVSWGY
 GCAQRRPGYVTVYVYVDMIKDTAANS"
 604699..605727
 /gene="TRYC"
 /note="putative"
 /number=1
 605510..647568
 /note="(vector pWE15A)"
 /organism="Homo sapiens"
 /cell_line="CgM1"
 /germline
 /sequenced_mol="DNA"
 /clone_lib="Kai Wang's"
 /citation=[1]
 605728..605887
 /gene="TRYC"
 /note="putative"
 /number=2
 605888..606945
 /gene="TRYC"
 /note="putative"
 /number=2
 606946..607199
 /gene="TRYC"
 /note="putative"
 /number=3
 607200..607594
 /gene="TRYC"
 /note="putative"
 /number=3
 607595..607731
 /gene="TRYC"
 /note="putative"
 /number=4
 607732..608030
 /gene="TRYC"
 /note="putative"
 /number=4
 608031..608181
 /partial
 /gene="TRYC"
 /note="putative"
 608214..608219
 /gene="TRYC"
 /note="putative"
 610541..610782

```

/rept_family="putative"
/rept_family="Alu"
complement(611096..611446)
/notes="putative"
/rept_family="LINE 1"
join(614744..614783,615820..615979,617038..617291,
617692..617828,618126..618275)
/genes="TRYD"
/notes="putative"
/pseudo
/product="trypsinogen D"
join(<614744..614783,615820..615979,617038..617291,
617692..617828,618126..618275)
/genes="TRYD"
/notes="putative"
<614744..614783
/genes="TRYD"
/notes="putative"
/pseudo
/number=1
614784..615819
/genes="TRYD"
/notes="putative"
/number=1
615820..615979
/genes="TRYD"
/notes="putative"
/pseudo
/number=2
615980..617037
/genes="TRYD"
/notes="no splice donor; putative; does not fit consensus"
/number=2
617038..617291
/genes="TRYD"
/notes="putative"
/pseudo
/number=3
617292..617691
/genes="TRYD"
/notes="putative"
/number=3
617692..617828
/genes="TRYD"
/notes="putative"
/pseudo
/number=4
617829..618125
/genes="TRYD"
/notes="putative"
/number=4
618126..618275
/genes="TRYD"
/notes="putative"
/pseudo
/number=5
618308..618313
/genes="TRYD"
/notes="putative"
complement(620651..620747)
/notes="putative"
/rept_family="Alu"
complement(620936..621210)
/notes="putative"

```

```

CDS
/rept_family="LINE 1"
join(625027..625066,626095..626254,627312..627565,
627965..628101,628402..628554)
/genes="TRYE"
/notes="putative; NCBI gi: 624714"
/codon_start=1
/product="trypsinogen E"
/translation="MNLILITFVAANAFAFDDDKIVGGYICEENSVPQVSLNSG
YHFCGSLISEQVSVSAGHCYKSRIQVRIGEHNTIEVLEGNQFINAAKIRHPKYNR
TLNDJLLIKLSSPAVINRSVSAISLPTAPPAAGTESLISGMGNTLSSGADYPDELQC
LDAPVLISQAECESYFGKITNNMFCVGFLEGKDSQCGSPVVSNGELQIVSWGY
GCAQKRRFGVYTKVYNYVDWIKDTIAANS"
join(<625027..625066,626095..626254,627312..627565,
627965..628101,628402..628554)
/genes="TRYE"
<625027..625066
/genes="TRYE"
/notes="putative"
/number=1
625067..626094
/genes="TRYE"
/notes="putative"
/number=1
626095..626254
/genes="TRYE"
/notes="putative"
/number=2
626255..627311
/genes="TRYE"
/notes="putative"
/number=2
627312..627565
/genes="TRYE"
/notes="putative"
/number=3
627566..627964
/genes="TRYE"
/notes="putative"
/number=3
627965..628101
/genes="TRYE"
/notes="putative"
/number=4
628102..628401
/genes="TRYE"
/notes="putative"
/number=4
628402..628590
/genes="TRYE"
/notes="putative"
/number=5
628591..628590
/genes="TRYE"
complement(630936..631210)
/notes="putative"
/rept_family="Alu"
complement(631452..631928)
/notes="putative"
/rept_family="LINE 1"
complement(635245..635510)
/notes="putative"
/rept_family="Alu"
complement(635729..635993)

```

```

repeat_unit /note="putative"
            /rpt_family="Alu"
            635948..636654
repeat_unit /note="putative"
            /rpt_family="HOMERSP1B"
            complement(636726..636988)
repeat_unit /note="putative"
            /rpt_family="Alu"
            complement(637950..638214)
misc_signal /note="putative"
            /rpt_family="Alu"
            640240..640267
            /gene="TCRBD1"
misc_signal /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            640240..640248
            /gene="TCRBD1"
misc_signal /note="RSS nonamer - awaiting approval of new feature key"
            640249..640260
            /gene="TCRBD1"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            640261..640267
            /gene="TCRBD1"
D_segment /note="RSS heptamer - awaiting approval of new feature
            key"
            640268..640279
            /gene="TCRBD1"
misc_signal /note="putative"
            640280..640318
            /gene="TCRBD1"
misc_signal /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            640280..640286
            /gene="TCRBD1"
misc_signal /note="RSS heptamer - awaiting approval of new feature
            key"
            640287..640309
            /gene="TCRBD1"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            640310..640318
            /gene="TCRBD1"
misc_signal /note="RSS nonamer - awaiting approval of new feature key"
            640307..640334
            /gene="TCRBJIS1"
misc_signal /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            640307..640315
            /gene="TCRBJIS1"
misc_signal /note="RSS nonamer - awaiting approval of new feature key"
            640316..640327
            /gene="TCRBJIS1"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            640328..640334
            /gene="TCRBJIS1"
            /note="RSS heptamer - awaiting approval of new feature
            key"
J_segment 640335..640382
            /gene="TCRBJIS1"
            /note="putative"
misc_signal 641044..641071
            /gene="TCRBJIS2"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            641044..641052

```

```

misc_signal /gene="TCRBJIS2"
            /note="RSS nonamer - awaiting approval of new feature key"
            641053..641064
            /gene="TCRBJIS2"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            641065..641071
            /gene="TCRBJIS2"
            /note="RSS heptamer - awaiting approval of new feature
            key"
J_segment 641072..641119
            /gene="TCRBJIS2"
            /note="putative"
misc_signal 641657..641684
            /gene="TCRBJIS3"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            641657..641665
            /gene="TCRBJIS3"
            /note="RSS nonamer - awaiting approval of new feature key;
            putative"
            641666..641677
            /gene="TCRBJIS3"
            /note="RSS spacer - awaiting approval of new feature key"
            641678..641684
            /gene="TCRBJIS3"
            /note="RSS heptamer - awaiting approval of new feature
            key; putative"
            641685..641734
            /gene="TCRBJIS3"
            /note="putative"
            642252..642279
            /gene="TCRBJIS4"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            642252..642260
            /gene="TCRBJIS4"
            /note="RSS nonamer - awaiting approval of new feature key"
            642261..642272
            /gene="TCRBJIS4"
            /note="RSS spacer - awaiting approval of new feature key"
            642273..642279
            /gene="TCRBJIS4"
            /note="RSS heptamer - awaiting approval of new feature
            key"
J_segment 642280..642330
            /gene="TCRBJIS4"
            /note="putative"
            642525..642552
            /gene="TCRBJIS5"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            642525..642533
            /gene="TCRBJIS5"
            /note="RSS nonamer - awaiting approval of new feature key"
            642534..642545
            /gene="TCRBJIS5"
            /note="RSS spacer - awaiting approval of new feature key"
            642546..642552
            /gene="TCRBJIS5"
            /note="RSS heptamer - awaiting approval of new feature
            key"
J_segment 642553..642602
            /gene="TCRBJIS5"
            /note="putative"

```

misc_signal 643015..643042
/gene="TCRBJ1S6"
/note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
643015..643023
/gene="TCRBJ1S6"
/note="RSS nonamer - awaiting approval of new feature key"
643024..643035
/gene="TCRBJ1S6"
/note="RSS spacer - awaiting approval of new feature key"
643036..643042
/gene="TCRBJ1S6"
/note="RSS heptamer - awaiting approval of new feature
key"
J_segment 643043..643095
/gene="TCRBJ1S6"
/note="putative"
repeat_unit complement(643863..644120)
/note="putative"
/rpt_family="Alu"
CDS join(645749..646135, 646577..646594, 646747..646853,
647176..647196)
/partial
/gene="TCRBC1"
/note="This CDS feature is included to show the
translation of the corresponding C region. Presently
translation qualifiers on C_region features are illegal.
NCBI gi: 624715"
/codon_start=3
/translation="DINKVPPPEVAFEPSEAFISHTQKATLVCLATCFPPDHVLSW
WNGKEVHGSVTDPPQPIKEQPALNDSRYCLSSRLVSAFTWPNRHFRCQVQYGL
SENDEWTDRAKPYVTQIVSNAEWGRADCGFTSVSYQQGVLSATILYEILIGKATLYAV
LVSAVLVMAWKRKDF"
C_region join(645749..646135, 646577..646594, 646747..646853,
647176..647196)
/gene="TCRBC1"
/note="putative"
/codon_start=3
exon 645749..646135
/gene="TCRBC1"
/note="putative"
intron 646136..646576
/gene="TCRBC1"
/note="putative"
exon 646577..646594
/gene="TCRBC1"
/note="putative"
intron 646595..646746
/gene="TCRBC1"
/note="putative"
exon 646747..646853
/gene="TCRBC1"
/note="putative"
intron 646854..647175
/gene="TCRBC1"
/note="putative"
exon 647176..647403
/gene="TCRBC1"

source 647326..650397
/note="The gap between cosmid clones CBG1 and C29 was
closed by the sequencing of a PCR reaction product."
/organism="Homo sapiens"
/germline
polyA_signal /sequenced_mol="DNA"
/clone="CBG1/C29gap"
/citation=[1]
647398..647403
/gene="TCRBC1"
/note="putative"
misc_signal 649732..649759
/gene="TCRBD2"
/note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
misc_signal 649732..649740
/gene="TCRBD2"
/note="RSS nonamer - awaiting approval of new feature key"
649741..649752
/gene="TCRBD2"
/note="RSS spacer - awaiting approval of new feature key"
649753..649759
/gene="TCRBD2"
/note="RSS heptamer - awaiting approval of new feature
key"
D_segment 649760..649771
/gene="TCRBD2"
/note="putative"
misc_signal 649776..649814
/gene="TCRBD2"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 649776..649782
/gene="TCRBD2"
/note="RSS heptamer - awaiting approval of new feature
key"
misc_signal 649783..649805
/gene="TCRBD2"
/note="RSS spacer - awaiting approval of new feature key"
649806..649814
/gene="TCRBD2"
/note="RSS nonamer - awaiting approval of new feature key"
650285..684973
/note="(vector pWE15A)"
/organism="Homo sapiens"
/cell_line="CCM1"
/germline
/sequenced_mol="DNA"
/clone="C29"
/clone_lib="Kai Wang's"
misc_signal /citation=[1]
650392..650419
/gene="TCRBJ2S1"
/note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
650392..650400
/gene="TCRBJ2S1"
/note="RSS nonamer - awaiting approval of new feature key"
650401..650412
/gene="TCRBJ2S1"
/note="RSS spacer - awaiting approval of new feature key"
650413..650419

```

/ gene="TCRBJ2S1"
/ note="RSS_heptamer - awaiting approval of new feature
key"
550420..650469
/ gene="TCRBJ2S1"
/ note="putative"
550587..650614
/ gene="TCRBJ2S2"
/ note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
550587..650595
/ gene="TCRBJ2S2"
/ note="RSS_nonamer - awaiting approval of new feature key"
550596..650607
/ gene="TCRBJ2S2"
/ note="RSS_spacer - awaiting approval of new feature key"
550608..650614
/ gene="TCRBJ2S2"
/ note="RSS_heptamer - awaiting approval of new feature
key"
550615..650665
/ gene="TCRBJ2S2"
/ note="putative"
550874..650901
/ gene="TCRBJ2S3"
/ note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
550874..650882
/ gene="TCRBJ2S3"
/ note="RSS_nonamer - awaiting approval of new feature key"
550883..650894
/ gene="TCRBJ2S3"
/ note="RSS_spacer - awaiting approval of new feature key"
550895..650901
/ gene="TCRBJ2S3"
/ note="RSS_heptamer - awaiting approval of new feature
key"
550902..650950
/ gene="TCRBJ2S3"
/ note="putative"
551025..651052
/ gene="TCRBJ2S4"
/ note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
551025..651033
/ gene="TCRBJ2S4"
/ note="RSS_nonamer - awaiting approval of new feature key"
551034..651045
/ gene="TCRBJ2S4"
/ note="RSS_spacer - awaiting approval of new feature key"
551046..651052
/ gene="TCRBJ2S4"
/ note="RSS_heptamer - awaiting approval of new feature
key"
551053..651102
/ gene="TCRBJ2S4"
/ note="putative"
551146..651173
/ gene="TCRBJ2S5"
/ note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
551146..651154
/ gene="TCRBJ2S5"
/ note="RSS_nonamer - awaiting approval of new feature key"

```

```

misc_signal
551155..651166
/ gene="TCRBJ2S5"
/ note="RSS_spacer - awaiting approval of new feature key"
551167..651173
/ gene="TCRBJ2S5"
/ note="RSS_heptamer - awaiting approval of new feature
key"
551174..651221
/ gene="TCRBJ2S5"
/ note="putative"
551266..651293
/ gene="TCRBJ2S6"
/ note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
551266..651274
/ gene="TCRBJ2S6"
/ note="RSS_nonamer - awaiting approval of new feature key"
551275..651286
/ gene="TCRBJ2S6"
/ note="RSS_spacer - awaiting approval of new feature key"
551287..651293
/ gene="TCRBJ2S6"
/ note="RSS_heptamer - awaiting approval of new feature
key"
551294..651346
/ gene="TCRBJ2S6"
/ note="putative"
551483..651510
/ gene="TCRBJ2S7"
/ note="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
551483..651491
/ gene="TCRBJ2S7"
/ note="RSS_nonamer - awaiting approval of new feature key"
551492..651503
/ gene="TCRBJ2S7"
/ note="RSS_spacer - awaiting approval of new feature key"
551504..651510
/ gene="TCRBJ2S7"
/ note="RSS_heptamer - awaiting approval of new feature
key"
551511..651557
/ gene="TCRBJ2S7"
/ note="putative"
552697..652814
/ note="putative"
/ rpt_family="MIR"
Join(655095..655481,655998..656015,656159..656265,
656557..656583)
/partial
/ gene="TCRCB2"
/ note="This CDS feature is included to show the
translation of the corresponding C_region. Presently
translation qualifiers on C_region features are illegal.
NCBI gi: 624716"
/ codon_start=3
/ translation="DLKNVFPPEVAFPESEAISSHQKATLVCLATGFPDHFVLSW
WNGKEVHSVSTDPOLKEOPALNDSRYCLSSRLVSAFQWPNRHFRCVOFYGL
SENDETQDRAKPVQTIVSAEWGRDCGFTSESYQQGVLSATILYELLGKATLYAV
LVSAIVLMAWVKRDSRG"
Join(655095..655481,655998..656015,656159..656265,
656557..656583)
/ gene="TCRCB2"
/ note="putative"
C_region

```

```

/codon_start=3
exon      655095..655481
          /gene="TCRBC2"
          /note="putative"
          /number=1
intron    655482..655997
          /gene="TCRBC2"
          /note="putative"
          /number=1
exon      655998..656015
          /gene="TCRBC2"
          /note="putative"
          /number=2
intron    656016..656158
          /gene="TCRBC2"
          /note="putative"
          /number=2
exon      656159..656265
          /gene="TCRBC2"
          /note="putative"
          /number=3
intron    656266..656556
          /gene="TCRBC2"
          /note="putative"
          /number=3
exon      656557..656774
          /gene="TCRBC2"
          /note="putative"
          /number=4
polyA_signal
          656769..656774
          /gene="TCRBC2"
          /note="putative"
          complement(658014..658405)
repeat_unit
          /note="putative"
          /rpt_family="MLT1B"
          complement(665732..665995)
repeat_unit
          /note="putative"
          /rpt_family="Alu"
          complement(666601..666639)
misc_signal
          /gene="TCRBC20S1"
          /note="RSS - awaiting approval of new feature key; nonamer
          is at 5' end; putative"
          666601..666609
          /note="RSS nonamer - awaiting approval of new feature key"
          666610..666632
          /note="RSS spacer - awaiting approval of new feature key"
          666633..666639
          /note="RSS heptamer - awaiting approval of new feature
          key"
          complement(join(666640..666931,667298..667340))
          /partial
          /gene="TCRBC20S1"
          /note="This CDS feature is included to show the
          translation of the corresponding V segment. Presently
          translation qualifiers on V_segment features are illegal.
          NCBI gi: 624717"
          /codon_start=1
          /translation="MCLSLALLCTFFGVRSGTHQWPATIVQVGPISLCTVEG
          TSNPNLWYRQAGRGQLLFYSGVIGQISSEVPQNLASRPQDRQFILSKKLLISD
          SGFYICAWSV"
          complement(join(666640..666931,667298..667340))
          /gene="TCRBC20S1"
          /note="putative"
          /codon_start=1
V_segment

```

```

exon      complement(666640..666931)
          /gene="TCRBC20S1"
          /note="putative"
          /number=2
intron    complement(666932..667297)
          /gene="TCRBC20S1"
          /note="putative"
          /number=1
exon      complement(667298..>667340)
          /gene="TCRBC20S1"
          /note="putative"
          /number=1
repeat_unit
          670197..670685
          /note="putative"
          /rpt_family="MLT1D"
          complement(671437..671825)
repeat_unit
          /note="putative"
          /rpt_family="MLT1D"
          complement(676821..679089)
repeat_unit
          /note="putative"
          /rpt_family="LINE 1"
          680649..680908
          /note="putative"
          /rpt_family="Alu"
          683515..683762
          /note="putative"
          /rpt_family="Alu"
          199337 a 146423 c 144088 g 195125 t
BASE COUNT
ORIGIN

DB 51; Score 26; Match 74.1%; QryMatch 2.4%; Pred. No. 7.09e-03;
Matches 40; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 322953 cccacctttccaccctcgcagtcacatccaccccccaacttcctgcacctt 323006
          ||||| ||||| | || ||||| ||||| ||||| |||||
Qy 931 CCCACCTTGGCCACCCCTGTGTGTCAGCTCCACCCCTGCTTCTGACCCCTTCT 984

Search completed: Thu Sep 28 00:16:46 1995
Job time : 624 secs.

```